

THIS WEEK

EDITORIALS

WORLD VIEW Europe should unite to lead the charge on open access **p.503**

SHRUNKEN SEAS Australian rain gobbles up world's missing water **p.504**



TUMMY RUMBLES Gut bacteria eavesdrop on neighbours to beat drugs **p.505**

Hidden heat

Scientists are homing in on the reasons for the current hiatus in global warming, but all must recognize that the long-term risk of warming from carbon dioxide remains high.

This week, *Nature* publishes a study online suggesting that a recent cooling trend in the tropical Pacific Ocean can explain the current hiatus in global warming. Authored by a pair of scientists at the Scripps Institution of Oceanography in La Jolla, California, the paper does not say why the Pacific seems to have entered a prolonged 'La Niña' phase, in which cooler surface waters gather in the eastern equatorial Pacific. It is also silent on where the missing heat is going. But it does suggest that this phenomenon — affecting as little as 8% of Earth's surface — could temporarily counteract the temperature increase expected from rising greenhouse-gas emissions (Y. Kosaka and S.-P. Xie *Nature* <http://dx.doi.org/10.1038/nature12534>; 2013).

Previous modelling studies have linked the pause to La-Niña-like conditions that have prevailed since 1999, suggesting that heat that would otherwise go into the atmosphere is getting buried deeper in the ocean. And scientists at the National Center for Atmospheric Research in Boulder, Colorado, have a study in the press indicating that decades in which global air temperature rises rapidly — including the 1980s and 1990s — are associated with warmer temperatures in the tropical Pacific, as exemplified by La Niña's opposite effect, El Niño (G. A. Meehl *et al. J. Climate* <http://doi.org/nkw>; 2013). The Scripps researchers also confirmed that El-Niño-like conditions can boost global temperatures.

Scientists seem to be homing in on an important lever in the climate system. And none too soon. Although a prolonged hiatus in warming does not necessarily contradict prevailing theory, this one came as a surprise and has been used to discredit the climate-science community. The story will probably not end there. Scientists know that the Sun has been in a prolonged solar minimum for several years, which means

less incoming energy, and there may yet be a role for sunlight-blocking aerosols — human pollution and volcanic ash — and other factors in the hiatus. But at least a better explanation of the climate system is beginning to take shape.

All of this comes as the Intergovernmental Panel on Climate Change (IPCC) prepares to release the first instalment of its fifth assessment report. The hiatus in warming is at the centre of an ongoing debate about 'equilibrium climate sensitivity', which is the amount of warming that would be expected over the long term owing to a doubling of atmospheric carbon dioxide levels. Several papers have assessed the most recent data and conclude that the climate may not be as sensitive to greenhouse gases as was previously thought. The latest draft of the IPCC summary for policy-makers accounts for this — just. It suggests a likely climate sensitivity of 1.5–4.5 °C, compared with a range of 2–4.5 °C in the IPCC's last assessment report.

Some argue that recent temperature trends show that the climate problem is less urgent. One can only hope that this is so, and scientists will continue to probe the matter. But policy-makers would be foolhardy to think that the danger has receded. Although scientists understand the basic physics, nobody can know how the numbers will turn out, as shown by the various temperature projections. Plenty of other lines of evidence, including palaeoclimate data and modern modelling experiments, support the higher end of these.

Ultimately, the decision over how to characterize climate sensitivity will fall to government officials who will approve — under the watchful eye of scientists — the latest IPCC documents in Stockholm next month. Whatever their decision, the underlying science has not changed. ■

Beyond compare

Metaphors are like cheese — often desirable but sometimes full of holes.

In an Amazon review of the 2013 book *Creation*, written by former *Nature* editor Adam Rutherford, one critic takes issue with what he describes as the work's "wordy trickery". "I lost count," the reviewer complains, "of the number of times that an obscure word or metaphor was helicoptered in."

Supplied by airlift or not, the online grumble illustrates both the irritation that some feel at analogies and metaphor in scientific writing, as well as the ease with which they can be, ahem, helicoptered in. It's a debate as old as, well, the hills. And one revived in a Comment article on page 523 that fizzles with ... [*That's enough metaphors — Ed.*].

In the piece, Eleonore Pauwels, a public-policy researcher in

the Science and Technology Innovation Program at the Woodrow Wilson International Center for Scholars in Washington DC, argues that biologists confuse the public when they borrow terms from engineering. Turning on genetic switches and assembling molecular components are processes that are more complex and ill-defined than the imagery might suggest, she says.

Metaphors, writers insist, breathe life into scientific language. But that is the problem, others say. The tendency to anthropomorphize the natural world was dubbed the "pathetic fallacy" by the nineteenth-century art critic John Ruskin. He would have hated the selfish gene.

Still, metaphor has a legitimate place in science. The idea that electrons orbit the nucleus like planets go around the Sun sets up testable hypotheses. Perhaps the problems come when scientific metaphors seek points for artistry rather than aiming for the quiet satisfaction of a job well done.

The English poet John Donne (1572–1631) famously compared the bound souls of lovers to a pair of compasses. "If they be two, they are two so / As stiff twin compasses are two." To compare a pair of compasses to the souls of lovers, however, would be wordy trickery too far. ■



A coordinated approach is key for open access

Cooperation and a clear set of aims are essential for Europe to be a front runner in making research freely available, says **Christoph Kratky**.

Even the most optimistic advocates of open access to academic publications must admit that we are years — and perhaps decades — away from full conversion to such a system. It is easy to call for open access, but more difficult to make it happen. More science funders must put their money where their mouths are, and back their positive words with action. It will not be cheap, but the longer we wait, the more expensive it will be.

Open access (OA) was an idea originally put forward by activists within the scientific community and later taken up by science policy-makers through declarations signed in Budapest, Bethesda in Maryland and Berlin. A decade after these statements, we are again faced with a wave of declarations on this topic, such as the position paper published in April by Science Europe, a research organization in Brussels, and the action plan prepared by the Global Research Council that was announced in May. All such declarations agree that the public that funds the research should have free access to the results, and that the current subscription-based publication model should be replaced.

Much progress has been made. The number of OA publications is on the rise, as is their reputation. Innovative OA concepts such as the Public Library of Science (PLOS) and BioMed Central have been established, some of which have proved economically viable. Countries such as the United Kingdom and disciplines such as high-energy physics have set their sights on moving to OA. And the European Commission plans to impose OA requirements on publications based on research funded under the Horizon 2020 Programme.

Yet despite this progress, a worrying imbalance remains between the efforts of research funders (including organizations that perform research), which can act only at a local level, and big publishing houses, which act globally. As a result, countries and institutions have different OA policies and behaviours that form a confusing patchwork. Some have explicit OA policies; others do not. Some require; others recommend. Some offer funds to pay for OA costs; others do not. Some have opted for 'gold' OA, which demands that publishers make papers freely available; others prefer 'green' OA, which allows researchers to archive the work.

Green OA seems a more workable solution, at least in the short term, but it is a mess. Many scientists simply don't know how, where and when to self-archive their papers. This confusing picture is confounded by the controversial question of whether widespread green OA can work as a cheap way to force reluctant publishers to adopt OA.

Some argue that it will succeed, because libraries will cancel subscriptions as soon as

enough self-archived articles are available. Others point out that archiving often happens only after an agreed embargo period — during which the only way to access the work is to pay the journal — and they say that subscriptions will endure. After all, who wants to settle for last week's newspapers? If publishers find it economically sustainable to establish a green world of subscription journals with embargo periods of six months or more, this road would turn out to be a dead end and thus fail to promote the desired transition to full OA.

The quality control imposed by publishers helps to ensure the integrity of the scholarly system, and that warrants financial compensation. Yet the current system of subscription journals is a classic example of a dysfunctional market, leading to high costs for the mostly publicly funded scholarly system. It is naive to expect that publishers might be persuaded to exchange a profitable business model for a potentially less profitable one.

Those who push for full OA must take firmer steps, and Europe should take the lead. Funding bodies should agree through their umbrella organizations on clear and uniform rules for the self-archiving of publications for both authors and users, with the support of appropriate legislation by the European Commission. Embargo periods of 6 to 12 months should be allowed, but only for the first few years.

Non-profit publishers need funds to move their high-ranking journals — which offer better value for money than commercial rivals — to OA. Funders should help to establish new, non-commercial academic publication models, which could be hosted by institutions such as universities, research organizations and learned societies.

This strategy could contribute to revitalizing the market, which is hindered by a lack of competition between few oligopolistic publishers.

It is easy to call for more money, but in recent years the Austrian Science Fund (FWF) in Vienna has provided it. The FWF has gradually increased its OA publishing costs, which now amount to about 0.8% (€1.5 million, US\$2 million) of the organization's annual budget. These costs are substantial compared with those of similar institutions. In a world of globalized research, however, national funding agencies quickly reach their limits. Only through close cooperation, starting at a European level, can we develop and implement models to accelerate the transition to full OA. It will be cheaper to invest now than to prolong the agony. ■

Christoph Kratky is a professor of structural biology at the University of Graz, Austria, and was president of the Austrian Science Fund from 2005 to 2013.
e-mail: christoph.kratky@uni-graz.at

**THOSE WHO PUSH
FOR FULL
OPEN ACCESS
MUST TAKE FIRMER
STEPS, AND
EUROPE
SHOULD TAKE
THE LEAD.**

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/pbilal

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

OBESITY

Heavy toll of stomach surgery

Weight-loss surgery improves the health of fat rats, but puts their offspring at risk of obesity and diabetes.

Randy Seeley and his team at the University of Cincinnati, Ohio, put female rats on a high-fat diet and then performed surgery to excise part of the rats' stomachs. After surgery, rats ate less food, lost weight and had fewer diabetes symptoms. Their offspring, however, were smaller and more susceptible to obesity and glucose intolerance compared with pups born to obese rats that had sham surgery.

For women, the authors say, weight loss alone may not stop them from passing on the effects of obesity to their children.

Science Transl. Med. 5, 199ra112 (2013)

MATERIALS

Electronic fabrics survive the wringer

Lightweight carbon sheets can be attached securely to nylon, cotton and polyester to create textiles that conduct electricity but don't fall apart in the wash.

Byung Hoon Kim at Incheon National University in South Korea and his colleagues smeared fibres with bovine serum albumin, a protein used in many biology labs. Albumin acted as a molecular glue; by



JONATHAN WOOD/GETTY

CLIMATE SCIENCES

Australia's record rains lowered sea level

Australia soaked up so much rain between early 2010 and late 2011 that global sea levels temporarily dropped.

John Fasullo of the US National Center for Atmospheric Research in Boulder, Colorado, and his colleagues used satellite measurements and tidal gauges to show that the mean global sea level — which had been rising by around 3 millimetres per year — fell by 7 mm. Using

measurements of Earth's gravity field, the team found a parallel increase in the mass of water on land, particularly in Australia, where topography prevents most water from flowing into the ocean.

Australia's record rainfall and flooding (pictured; in Queensland) probably came from a triad of regional climate patterns, including La Niña surface cooling in the tropical Pacific. *Geophys. Res. Lett.* <http://doi.org/ngx> (2013)

changing the charge on the fibres, it attracted sheets of graphene oxide (single-atom-thick layers of carbon atoms, topped with oxygen). The modified threads remained flexible (**pictured**) and could carry a current after being bent, washed and subjected to temperatures between -53°C and 52°C .

Adv. Mater. <http://doi.org/njx> (2013)

CANCER

Revived genes give therapy a boost

A drug that inhibits DNA-modifying enzymes makes stubborn lymphoma cells sensitive to chemotherapy.

A team led by Leandro Cerchietti and Ari Melnick of Weill Cornell Medical College in New York studied drugs that prevent DNA methylation, a modification that 'silences' genes. Work in cell lines and mouse models for diffuse large B-cell lymphoma (DLBCL) showed that such a drug reactivated genes that made the anti-cancer agent doxorubicin more effective. Twelve patients with DLBCL were given an anti-methylation drug and standard chemotherapy; 11 responded. Biopsies showed decreased methylation and increased activity of a gene called *SMAD1*, which is often silenced in resistant DLBCL. The results suggest

that targeting excessive DNA methylation could reverse treatment resistance.

Cancer Discov. 3, 1–18 (2013)

ENERGY STORAGE

Power surge for flow batteries

Expensive membranes can be eliminated from flow batteries without compromising performance.

Flow batteries, used to store power in electricity grids, rely on fuels pumped through a power-generating region bounded by electrodes. Membranes typically stop fuels from reaching areas that would cause a short circuit. Some cheaper batteries are

YONG JU YUN

built so fuel flows smoothly and no membrane is required, but these have low power density and are usually not rechargeable.

Cullen Buie and his colleagues at the Massachusetts Institute of Technology in Cambridge engineered a membraneless hydrogen-bromine battery that can pump fuels at very high concentrations. The design boosts power density to three times that reported for other membrane-free batteries, and the battery can be recharged. *Nature Commun.* <http://doi.org/njt> (2013)

ATOMIC PHYSICS

Super-stable clocks

The most stable atomic clock built so far loses or gains less than one second every quintillion seconds (equal to 10^{18} seconds, or more than 30 billion years), which is about ten times better than the previous record.

Andrew Ludlow at the National Institute of Standards and Technology in Boulder, Colorado, and his colleagues used lasers to monitor the oscillation of energy levels in two sets of ultracold ytterbium atoms that had been trapped using magnetic fields and laser beams. The clocks ticked in sync with a stability of about one part in 10^{18} .

Such stable clocks could benefit satellite communication and navigation, and would enable space-based tests of general relativity to be conducted about 1,000 times more precisely than with clocks available today.

Science <http://doi.org/nj9> (2013)

NEUROSCIENCE

Single gene tweak for prion diseases

Engineered mouse strains develop rare prion diseases without the need for an infectious agent.

Susan Lindquist of the

Whitehead Institute for Biomedical Research in Cambridge, Massachusetts, and her colleagues have created mouse models for two fatal human diseases caused by prions — misfolded proteins that spur other proteins to misfold. Creutzfeldt–Jakob disease and fatal familial insomnia affect different parts of the brain and arise from different mutations in the same gene. Rather than inserting mutated genes into the mouse genome — a tactic that ineffectively mimics the disease — the researchers mutated the prion gene in place.

Single amino-acid changes in the mouse prion protein caused the brain degeneration that is characteristic of each disease, and both models produce prions that can infect other mice.

Proc. Natl Acad. Sci. USA <http://doi.org/njv> (2013)

CANCER BIOLOGY

Cancer mutation has a long reach

Some DNA mutations that promote cancer may work through extra copies not of genes, but of genetic regulatory sequences.

Duplications of the chromosomal sequences 17q23 and 20q13 predict poorer outcomes in certain breast cancers — but no genes in these regions suggest a mechanism for how. Tim Huang and co-workers at the University of Texas Health Science Center in San Antonio sequenced genomic regions that physically interact with other stretches of DNA that bind to the oestrogen receptor, a protein complex implicated in breast cancer. This identified sites within 17q23 and 20q13 called distant oestrogen response elements (DEREs), which help the genetic material to form loops that interact with far-off genomic regions. Extra DEREs promoted more looping and deregulated genes in ways expected to promote tumour growth.

Cancer Cell 24, 197–212 (2013)

COMMUNITY CHOICE

The most viewed papers in science

ANIMAL BEHAVIOUR

Whales hear the noise

HIGHLY READ
on rspb.royalsocietypublishing.org in July

Simulated military sonar can cause blue whales (*Balaenoptera musculus*) to stop eating and swim away.

Jeremy Goldbogen at Cascadia Research Collective in Olympia, Washington, and Brandon Southall at Southall Environmental Associates in Aptos, California, and their colleagues attached sensors to 17 blue whales off southern California. The team tracked the animals before, during and after playing sonar-like sounds or white noise from a research vessel. Whales at the surface typically did not respond, but whales at depth often showed behaviours such as swimming faster or ceasing to feed.

Human-produced noises could keep baleen whales from foraging and so affect their fitness, the authors say. *Proc. R. Soc. B* 280, 20130657 (2013)



MICROBIOLOGY

Gut pathogen spies others' signals

A disease-causing bacterium 'eavesdrops' on another, usually harmless, bacterial species to help it survive antibiotics.

Under certain conditions, populations of *Escherichia coli* that live in the gut secrete indole, a signalling molecule that makes them more tolerant of antibiotics.

Although the gut pathogen *Salmonella typhimurium* (pictured) cannot produce indole, it still responds to it. James Collins at Boston University, Massachusetts, and his colleagues showed that *S. typhimurium* exposed to indole or to indole-producing *E. coli* survived better under antibiotics than did bacteria in indole-free conditions. In both species, indole strongly boosted the expression of stress-response genes known to help bacteria withstand drugs and assaults by the human immune system.

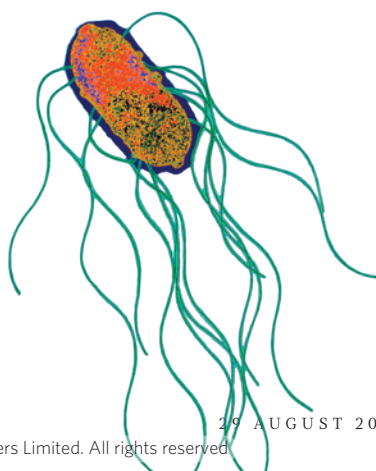
Such interactions between species could help harmful bacterial infections to persist, the authors say.

Proc. Natl Acad. Sci. USA <http://doi.org/njw> (2013)

NATURE.COM

For the latest research published by Nature visit:

www.nature.com/latestresearch



SEVEN DAYS

The news in brief

POLICY

Fishing fallout

The European Union (EU) has escalated a trade war with the Faroe Islands over fishing in the Atlantic. On 20 August, it adopted measures that will ban the importation of herring and mackerel fished by the Faroes from Atlantic and Scandinavian waters. Some boats from the Faroes will also be banned from European ports. The Faroe Islands has condemned the economic sanctions, which the EU says are a response to the country's continued overfishing of herring. See go.nature.com/qmuwst for more.

RESEARCH

Bat signal

Bats have been implicated in the spread of the deadly Middle East respiratory syndrome coronavirus (MERS-CoV), which has killed 48 people since April 2012. On 21 August, researchers reported that of 96 bats surveyed in Saudi Arabia, one animal carried a small viral genetic fragment identical to part of the MERS-CoV genome. The scientists say that the match connects bats to the spread of the virus, but critics argue that the fragment is too small to determine whether the animal had been infected with MERS-CoV, or a closely related virus. See go.nature.com/b6u7zr for more.

X Prize cancelled

The X Prize Foundation has cancelled a contest that would have awarded US\$10 million to the first team to produce 100 high-quality genome sequences in 30 days for \$1,000 apiece. In a 22 August blog post, Peter Diamandis, chair of the X Prize Foundation, said that the contest was made



TERRY ZAPEROCH/NASA

More NASA surplus up for grabs

Three huge transport platforms for hauling space shuttles have outlived their purpose at NASA's Kennedy Space Center in Cape Canaveral, Florida. The space agency announced on 20 August that it is "seeking concepts" for the future use of the 3,700-tonne, 50-metre-long steel structures (pictured, beneath the space shuttle *Atlantis*), which carried Saturn rockets and space shuttles from

where they were assembled to their launch pads. NASA has already put one of its shuttle launch pads up for lease at the facility, attracting interest from commercial space companies SpaceX in Hawthorne, California, and Blue Origin in Kent, Washington. NASA says that the mobile launch platforms could be of use in future commercial launches at the Kennedy Space Center.

redundant by the tumbling cost of genetic sequencing — a factor that contributed to a lacklustre showing of only two competitors (see *Nature* **497**, 535 and 546–547; 2013).

PEOPLE

Forensics fiasco

The number of people whose criminal cases were potentially affected by Annie Dookhan, a chemist alleged to have falsified test results at a Massachusetts crime laboratory, now exceeds 40,000, according to a state-commissioned review completed on 20 August. A database is being assembled

to aid prosecutors, courts and defence lawyers as they review the cases whose evidence came into contact with Dookhan — a list that now includes nearly 3,000 more individuals than previously thought. Dookhan pleaded not guilty to charges of obstruction of justice and perjury (see *Nature* **490**, 153–154; 2012).

FACILITIES

Atomic hopes

Japan's high-energy physics community selected on 23 August a possible site for the International Linear Collider, a proposed atom smasher designed to make

precision measurements of known particles. Located in the Kitakami mountains of the Tohoku region, the site could be eligible for reconstruction funds set aside after the 2011 earthquake and tsunami that struck the region. The researchers hope to spur the government of Japan to put in an official bid for the machine, which is expected to cost US\$8 billion to construct.

Telescope strike

Less than six months after the inauguration of a US\$1.4-billion radio telescope array in Chile, astronomical observations have ceased because of a labour strike.

Work at the Atacama Large Millimeter/submillimeter Array (ALMA) ground to a halt on 22 August after contract negotiations broke down with the local workers' union, which is demanding pay rises, bonuses and shorter shifts. Most staff astronomers have left the site, but many of the 195 striking employees — who are mainly Chilean and include engineers, data managers and administrative staff — remain at the facility. See go.nature.com/vhxfrv for more.

WISE up again

NASA will revive the dormant Wide-field Infrared Survey Explorer (WISE) to help search for asteroids that the agency could grab and relocate near the Moon for study. The spacecraft will be switched back on in September, NASA announced on 21 August. The probe had been turned off in early 2011 after finishing its primary mission, but NASA is looking for all the help it can get in finding a small asteroid, in just the right orbit, for its planned asteroid-retrieval mission (see *Nature* **499**, 261–262; 2013).

Telescope deal

The Green Bank radio telescope in West Virginia will receive a US\$1-million boost from West Virginia University in Morgantown, according to an agreement announced on

23 August. The university will provide the funds to the facility over the next two years in exchange for a dedicated share of observation time. The US National Science Foundation, which owns the telescope, received recommendations last year to shed Green Bank and other observatories from its portfolio by 2017 to make way for new telescopes (see *Nature* **488**, 440; 2012).

BUSINESS

Biotech buyout

Biotech giant Amgen will buy Onyx Pharmaceuticals for about US\$10.4 billion, the firm, based in Thousand Oaks, California, announced on 25 August. The deal is one of the five largest-ever takeovers of a biotech firm. Onyx, based in South San Francisco, California, has focused on cancer therapies, including the bone-marrow cancer drug carfilzomib (Kyprolis), which was approved last year in the United States.

EVENTS

Old globe

Cartographers may have identified the oldest known globe depicting the New World, according to a study published last week (*S. Missinne The Portolan* issue 87, 8–24; 2013). The engraved, grapefruit-sized



orb (pictured), which is made of fused ostrich egg halves, depicts geographical details from early European explorers in a style reminiscent of Leonardo da Vinci. Carbon dating and other analyses date it to 1504. The copper Lenox Globe, dated to 1510 and previously considered to be the oldest showing the New World, may in fact be modelled on this globe, which was discovered in a 2012 London map fair.

Fukushima leak

Following the leakage of some 300,000 litres of radioactive cooling water at the Fukushima nuclear plant, Japan's Nuclear Regulation Authority says that it is considering upgrading the official severity level from 1 (an 'anomaly') to 3 (a 'serious incident'). The radioactive water, which came from a steel storage tank, is thought to have seeped into the ground about 50 metres above sea level. Observers worry that

COMING UP

5–7 SEPTEMBER

Scientists and engineers meet in Athens for the 13th International Conference on Environmental Science and Technology, to discuss topics such as emerging pollutants, soil remediation and water-resources management.

go.nature.com/s4s68s

groundwater is likely to carry the contaminated water into the nearby ocean.

Wildfires rage

Some 40 large wildfires are burning in the western United States, including a blaze that crossed into Yosemite National Park in California on 23 August, according to the US Forest Service. The quickly spreading blaze, one of a dozen major fires burning in the state, prompted California governor Jerry Brown to declare a state of emergency on 22 August. The US government has spent more than US\$1.2 billion this year to fight wildfires across the country, says the National Interagency Fire Center in Boise, Idaho.

Flood aftermath

Heavy rains and flooding in Sudan have killed 48 people, injured 70 more and placed an estimated 320,000 people in danger, the World Health Organization (WHO) said on 21 August. Agency officials are concerned about a mounting disease risk caused by sanitation failures in the wake of the flooding. Nearly 53,000 latrines have collapsed since heavy rains began in early August. The number of malaria cases in two affected regions has also risen over the past two weeks, the WHO reported.

NATURE.COM

For daily news updates see:

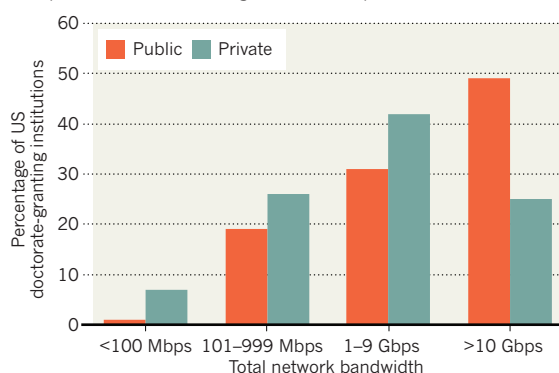
www.nature.com/news

TREND WATCH

Access to high-speed Internet has increased rapidly at US academic institutions over the past few years. The National Science Foundation estimates that in financial year 2012, 45% of all academic institutions could transmit data over the network at a rate of 2.5 billion bits per second or more, compared with just 6% in 2005. Among doctorate-granting institutions, public ones were more likely than their private counterparts to have access to the highest bandwidths.

UNEVEN NETWORK ACCESS AT US UNIVERSITIES

As availability of high-speed Internet spreads, public universities beat private schools at the highest network speeds.



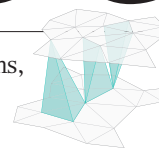
*Number of institutions: 260 public, 119 private; Gbps = gigabits per second; Mbps = megabits per second.

NEWS IN FOCUS

PUBLISHING Brazilian journals outed in scheme to ramp up impact factors **p.510**

WEATHER Temporary reprieve for Arctic as summer cyclones bolster sea ice **p.512**

BUSINESS As biotech booms, analysts worry if there will be a bust **p.513**



PHYSICS Wild ideas about the origins of space and time **p.516**

SSC



Launched in 2010, the Swedish satellite PRISMA has been testing a greener alternative to the toxic propellant hydrazine.

SPACE SCIENCE

Green fuels blast off

Propellants offer satellites greater efficiency and lower toxicity than liquid hydrazine.

BY ALEXANDRA WITZE

It looks like chardonnay, smells like glass cleaner and packs enough punch to shift a satellite. It is a Swedish-made 'green propellant', one that is fast becoming a viable fuel for manoeuvring craft in orbit. Along with a US-made propellant, it is providing an attractive alternative to hydrazine, the toxic chemical that has dominated this corner of the space industry for decades.

The Swedish propellant is currently nudging a satellite around in space, and may be used in a constellation of small commercial Earth-imaging satellites. The US-made fuel will take centre stage in a NASA test mission launching in 2015.

Both fuels offer higher efficiency, lower toxicity and are easier to handle than hydrazine,

which means that they can be loaded into spacecraft faster and more cheaply because launch-pad workers don't have to wear cumbersome full-body protection. Such a propellant "isn't green in the sense that it's totally environmentally friendly", says James Reuther, deputy associate administrator of NASA's space technology mission directorate. "But it's a heck of a lot easier to work with."

Green propellants probably won't completely replace hydrazine, the workhorse of many research satellites and interplanetary missions. Nor will they replace the powerful fuels typically used to launch rockets. But the new fuels, some of which are more tolerant of low temperatures, could enable cheaper and more flexible mission designs.

Hydrazine (N_2H_4) has been powering rocket

engines since the Second World War. It tends not to be burned like petrol; instead, a catalyst is used to trigger hydrazine's decomposition into ammonia, nitrogen and hydrogen, a process that releases chemical energy. The fuel is valued for being a stable liquid that can provide precision thrusts for small orbital adjustments.

But hydrazine also causes a host of health problems if breathed in or touched. The US National Toxicology Program classifies it as a probable human carcinogen. When the space shuttle *Columbia* disintegrated on re-entry in 2003, showering debris across Texas, Louisiana and other southern states, NASA warned people not to approach or touch the wreckage in part because of the risk of hydrazine exposure.

In 1995, Sweden's national space board funded work to explore alternatives that ►

► would perform at least as well as hydrazine but that would be easier to handle. A less unwieldy propellant would help to save time and money at launch pads, says Mathias Persson, president of ECAPS, the company near Stockholm that developed the green propellant.

Called LMP-103S, the Swedish fuel is based on ammonium dinitramide, a high-energy salt. It made its debut in 2010 aboard PRISMA, a Swedish satellite meant to demonstrate the fuel in precision flying with small thrusters. For comparison, the mission also carried hydrazine. It took 3 people 7 days to load the green propellant on the launch pad, and 5 people and 14 days to load the hydrazine.

ECAPS is now seeking general approval from the European Space Agency for the propellant. France's space agency, CNES, is considering it for a new line of small satellites, and the company Skybox Imaging of Mountain View, California, will begin using the fuel when it launches the third of its 24-craft Earth-imaging constellation in 2015. "We really believe this is going to be the future, especially for small spacecraft," says Jonny Dyer, Skybox's chief engineer.

Green fuels aren't just easier to handle; they can also provide better performance per kilogram of propellant than hydrazine (see 'Designer fuels'). That means more manoeuvres can be done on a single tank of fuel. In fact, the US green propellant came about when two Air Force research programmes collided: one looking at next-generation propulsion technologies,

DESIGNER FUELS

Hydrazine could be unseated as the go-to fuel for satellite orbital manoeuvres by two propellants with lower toxicity and more punch.

	Hydrazine	LMP-103S	AF-M315E
Developed by	Multiple	ECAPS, Sweden	US Air Force Research Laboratory
Contains	Hydrogen and nitrogen	Ammonium dinitramide	Hydroxyl ammonium nitrate
Specific impulse	2,373 N s kg ⁻¹	2,477 N s kg ⁻¹	2,609 N s kg ⁻¹
Density	1.01 g cm ⁻³	1.24 g cm ⁻³	1.465 g cm ⁻³
Toxicity (lethal dose in 50% of rats, LD ₅₀)	60 mg kg ⁻¹	1,300 mg kg ⁻¹	550 mg kg ⁻¹

Source: AFRL; ECAPS

the other studying extremely energy-dense materials.

The result was AF-M315E, a salmon-coloured liquid created by Tom Hawkins, a chemist at the Air Force Research Laboratory at Edwards Air Force Base in California. The fuel is based on hydroxyl ammonium nitrate, which is slightly more efficient than the Swedish propellant. "What I get excited about is the performance part of it," says Christopher McLean, the programme manager at Ball Aerospace & Technologies in Boulder, Colorado.

Ball is building NASA's Green Propellant Infusion Mission, a US\$45-million spacecraft expected to launch in 2015, carrying 14.2 kilograms of the Air Force fuel. Its five engines will burn in different operations for months, testing how reliably the engines thrust.

If it works, the new propellant could enable

future missions that are expensive to do at the moment, says McLean. AF-M315E undergoes a glass transition at -80°C, from which it can be heated back up with no change in its properties. That could make it more attractive than hydrazine for missions to ultra-cold locations, like a comet's surface or the Martian polar cap, as hydrazine has to be kept above its freezing point at all times, which wastes energy.

Hydrazine won't be displaced overnight. There is a long heritage of tanks and thrusters based around the chemical, and spacecraft engineers are accustomed to using those parts. The new propellants, says Hawkins, will have to demonstrate reliability and performance over time.

But change is coming, says Kjell Anflo, chief engineer for ECAPS. "We are just at the beginning." ■

PUBLISHING

Brazilian citation scheme outed

Thomson Reuters suspends journals from its rankings for 'citation stacking'.

BY RICHARD VAN NOORDEN

Mauricio Rocha-e-Silva thought that he had spotted an easy way to raise the profiles of Brazilian journals. From 2009, he and several other editors published articles containing hundreds of references to papers in each others' journals — in order, he says, to elevate the journals' impact factors.

Because each article avoided citing papers published by its own journal, the agreement flew under the radar of analyses that spot extremes in self-citation — until 19 June, when the pattern was discovered. Thomson Reuters, the firm that calculates and publishes the impact factor, revealed that it had designed a program to spot concentrated bursts of citations from one journal to another, a practice that it has dubbed 'citation stacking'. Four Brazilian journals were among 14 to have their impact factors suspended for a year for such

stacking. And in July, Rocha-e-Silva was fired from his position as editor of one of them, the journal *Clinics*, based in São Paulo.

"We've been caught wrong-footed," says Rocha-e-Silva, a retired physiologist. The editors of the other three Brazilian journals colored by Thomson Reuters remain in place. In addition to these four journals, "there are a few others which played a part in this game, and they escaped", he says.

Editors have tried before to artificially boost impact factors, usually by encouraging the citation of a journal's own papers. Each year, Thomson Reuters detects and cracks down on excessive self-citation. This year alone, it red-flagged 23 more journals for the wearily familiar practice. But the revelation that journals have gained excessively from citations elsewhere suggests that some editors may be searching for less detectable ways to boost their journals' profiles. In some cases, authors may

be responsible for stacking, perhaps trying to boost citations of their own papers.

The journals flagged by the new algorithm extend beyond Brazil — but only in that case has an explanation for the results emerged. Rocha-e-Silva says the agreement grew out of frustration with his country's fixation on impact factor. In Brazil, an agency in the education ministry, called CAPES, evaluates graduate programmes in part by the impact factors of the journals in which students publish research. As emerging Brazilian journals are in the lowest ranks, few graduates want to publish in them. This vicious cycle, in his view, prevents local journals improving.

Abel Packer, who coordinates Brazil's system of free government-sponsored journals, known as SciELO, says that the citation-stacking venture was "unfortunate and unacceptable". But he adds that many editors have long been similarly critical of the CAPES policy

because it encourages local researchers to publish in high-impact journals, increasing the temptation for editors to artificially boost their own impact factors, he says.

Brazilian editors have campaigned for years for CAPES to change its system. "But they have always adamantly refused to do this," Rocha-e-Silva says. Bruno Caramelli, editor of *Revista da Associação Médica Brasileira*, which was also hauled up by Thomson Reuters, says that by 2009, editors of eight Brazilian journals decided to take measures into their own hands (see 'Citation stacking').

But other editors involved in the agreement dispute Rocha-e-Silva's assertion that they aimed solely to increase impact factors; Caramelli and Carlos Roberto Ribeiro de Carvalho, editor of the *Jornal Brasileiro de Pneumologia*, argue that the idea was also to show off articles in Brazilian journals, attracting better contributions and raising quality all round.

The punishment for the four journals caught out has been severe: CAPES says that any articles they published in 2010–12 will not count in its October 2013 evaluation of graduate programmes. The journals have also been suspended from CAPES' evaluation system until their impact factors are restored.

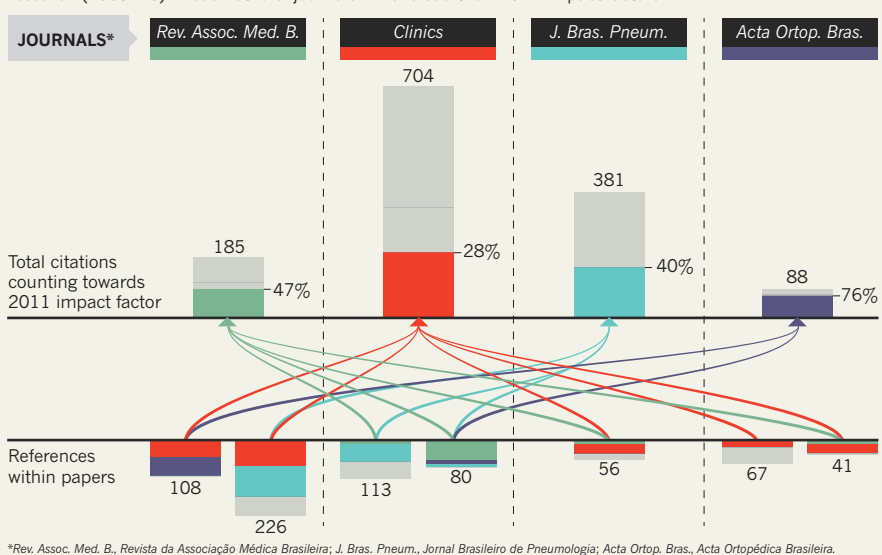
Citation stacking first came to light in 2012, after Thomson Reuters was tipped-off about a case in which (now-retracted) articles in two journals contained hundreds of references to a third, boosting its impact factor (see go.nature.com/rehuit). In response, the company pioneered its new algorithm, designed to spot the practice.

What happened in the cases of the other ten journals censured for citation stacking is unclear. One involves a close pattern of citations between three Italian journals (*International Journal of Immunopathology and Pharmacology*, *Journal of Biological Regulators and Homeostatic Agents* and *European Journal of Inflammation*) all with the same editor-in-chief, Pio Conti, an immunologist at the University of Chieti-Pescara. Many of the authors on the relevant papers are also at that university. Conti told *Nature* that he regretted that the anomalous citations had occurred, but that "we have no quick explanation of the patterns".

In another case, review articles with hundreds of references to *Science China Life Sciences* were meant not to lift its impact factor, but to clarify confusions after a rebranding and to "promote

CITATION STACKING

In 2011, four Brazilian journals published seven review papers with hundreds of references to previous research (2009–10) in each others' journals. This raised their 2011 impact factors.



the newly reformed journal to potential new readers", says its editor-in-chief Dacheng Wang, who is based at the Chinese Academy of Sciences' Institute of Biophysics in Beijing.

In a further case, the *Journal of Instrumentation* saw hundreds of cross-citations from papers authored in *SPIE Proceedings* by Ryszard Romaniuk, an electronic engineer who was part of the collaboration that put together the CMS experiment in the Large Hadron Collider at CERN, Europe's high-energy physics laboratory near Geneva in Switzerland. Romaniuk, who repeatedly cited CMS engineering papers, has not replied to *Nature*, nor to the *Journal of Instrumentation's* publishers — the Institute of Physics and the International School of Advanced Studies (SISSA).

Marie McVeigh, who leads *Journal Citation Reports*, Thomson Reuters' annual report of journal citation patterns and impact factors, says that the firm does not assume intent in the patterns it sees. "We just analyse data, and see where the impact factor, due to this unusual concentration of citations, is not a reflection of the journal's citation in the broader literature. Until we can algorithmically measure motive, we are just going to measure citations." She says that because citation stacking is only a problem if it excessively distorts journal rank, only

four Brazilian journals, and not all of those that Rocha-e-Silva says participated in the arrangement, have had their impact factors suspended.

Thomson Reuters has received some appeals against its rulings, with some editors arguing that to suspend impact factors for incoming citations of which they were unaware was unfair; or that outgoing citations did not affect their own journal. The firm will deal with these by September, McVeigh says. Rocha-e-Silva appealed against the suspension of *Clinics*, but his request was turned down.

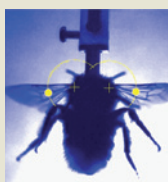
He feels that journal impact-factor scores are a bad way to judge scientists' work and that the real problem is CAPES' policy. But he also accepts responsibility for what he did. "We're not going to play that again wherever I become editor, if I ever do. I've burned my fingers once," he says.

The journals currently suspended for either self-citation or citation stacking represent only 0.6% of the 10,853 in Thomson Reuters' respected directory. But impact-factor anomalies will continue, McVeigh says. Aside from some journals with excessive self-citations, "we've already identified some journals citation stacking for next year", she says. ■

Additional reporting by Jeff Tollefson.



VIDEO OF THE WEEK



Insect wing muscles are not radically different than vertebrate muscles go.nature.com/h5gsi6

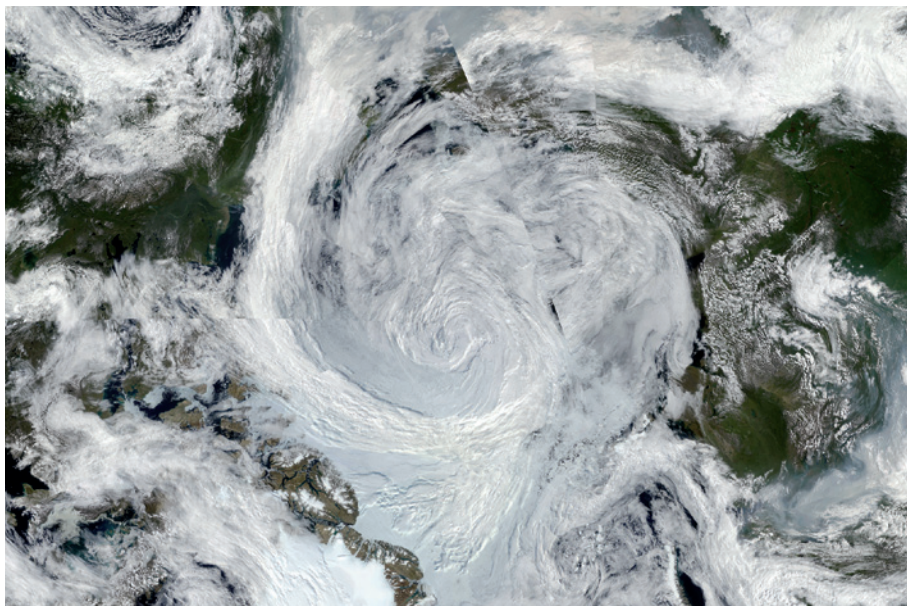
MORE NEWS

- Neolithic Europeans spiced their food go.nature.com/yzig4r
- Magnetic bubbles suggest Voyager 1 has left Solar System go.nature.com/7kclua
- Network theory suggests geography makes US electrical grid inherently unstable go.nature.com/woydyo

NATURE PODCAST



Global temperatures; home-grown brains; and outlandish ideas about space-time nature.com/nature/podcast



The Great Arctic Cyclone of 2012 chewed up weak ice in a way that could become the norm.

CLIMATE CHANGE

Summer storms bolster Arctic ice

This year's cyclones grant sea ice a temporary reprieve.

BY LAUREN MORELLO

When it comes to maintaining sea ice in the Arctic, it seems, there are good cyclones and there are bad ones.

Last year brought the bad kind. In early August 2012, an unusually large and intense storm barrelled into the Arctic from Siberia. The cyclone churned for nearly two weeks at the height of the summer melt season, contributing to the break-up of the ice, which was already thin and weak, owing to a warm, early start to the season. By September, Arctic sea ice had shrunk to a record low of just 3.4 million square kilometres.

Yet this year, the summer storms that have passed over the Arctic have fulfilled their usual role, bringing snow and cooler air to slow the melting. At present, with the summer thaw set to end in just a few weeks, sea-ice cover stands at around 5.7 million square kilometres (see 'Slippery slope').

The vast difference between the two melt seasons has researchers investigating the ways in which cyclones can either exacerbate or dampen the effects of climate warming on sea ice. Yet, as global warming continues and cyclones become more intense, many

researchers worry that summers like that of 2012 could become the norm.

Ian Simmonds, a climatologist at the University of Melbourne, Australia, says that there is no doubt that the 2012 cyclone was extreme. His calculations show that the storm, called the Great Arctic Cyclone, was the most intense August storm to hit the region since satellite monitoring began in 1979, and the 13th most

intense among all storms recorded there. In another study, researchers at NASA found that wind and wave action driven by the tempest caused a massive chunk of ice — covering an area of 400,000 square kilometres — to separate from the main pack (C. L. Parkinson & J. C. Comiso *Geophys. Res. Lett.* **40**, 1356–1361; 2013). The sheared-off portion eventually melted, depleting the overall ice cover and leaving the main pack more vulnerable to erosion from wind, waves and warm water stirred up by the storm.

Other research suggests that the storm's impact was not quite so severe. Axel Schweiger, a climatologist at the University of Washington in Seattle, ran a computer model to isolate the effects of the cyclone and found that, on its own, it was responsible for only about 150,000 square kilometres of the season's melt (J. Zhang *et al. Geophys. Res. Lett.* **40**, 720–726; 2013).

In contrast to last year, this summer's cyclones have bolstered the ice pack by bringing in cold air. Snow from the storms falls on the ice and nearby land, increasing the amount of sunlight that is reflected back towards space and further helping to keep temperatures down.

Yet scientists say that future summer storms are more likely to accelerate ice loss than to slow it, as rising temperatures continue to deplete the Arctic's ice cover. Thick, 'multi-year' ice that has survived more than one summer thaw once covered large expanses of the Arctic Ocean. Now those waters are dominated by thin, 'first-year' ice — which forms in autumn and melts the next summer, and is more vulnerable to the fierce winds and waves that the cyclones bring. "As the ice gets thinner and more dispersed, it's more of a sitting duck as storms come through," says Simmonds.

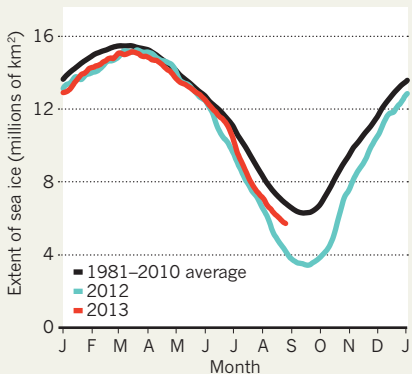
The disappearance of the ice may itself lead to more intense storms, says Claire Parkinson, a climatologist at NASA's Goddard Spaceflight Center in Greenbelt, Maryland. Disappearing ice leaves stretches of open ocean, allowing more warm water to get up into the atmosphere and fuel storms, she says. "The system is definitely interconnected."

So far, although some research suggests that Arctic cyclones may be growing more intense, several studies have found no apparent change in the frequency of late spring or summer storms there. "Whether we should expect more of these things, that is an open question," says Schweiger.

This year's smaller summer melt in the Arctic cannot mask the ongoing decline of the sea ice, says Julianne Stroeve, a climate scientist at the US National Snow and Ice Data Center in Boulder, Colorado. Seven of the eight lowest recordings of sea-ice extent occurred in the past eight years, she says. By early August this year, sea-ice cover was already lower than annual lows recorded in the 1970s and 1980s, with weeks of melting still to go. ■

SLIPPERY SLOPE

The 2013 sea-ice minimum will not reach the record low set in 2012, when a major cyclone added to losses.



SOURCE: NSIDC

BUSINESS

Biotech boom prompts fears of bust

A spate of successful public offerings has raised ample funding for biotechnology — and anxiety about when the good times will end.

BY HEIDI LEDFORD

It has been the summer of biotech. After years of fretting that investors had soured on the high-risk industry, untested biotechnology companies are all of a sudden going public. This year, 24 US firms have issued initial public offerings (IPOs), pumping US\$1.8 billion into the industry. Their stocks rose an average of 20% on the first day of trading. Another eight companies plan to follow suit in the coming months.

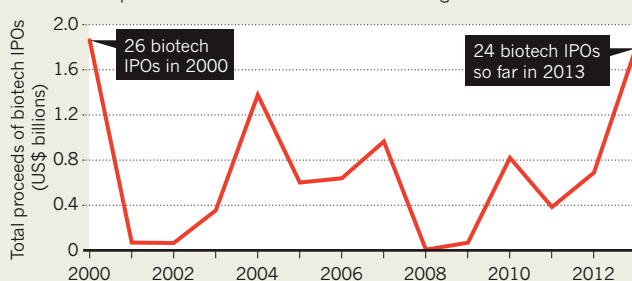
If they do, it will be a record-setting year for biotechnology (see 'Bullish on biotech'). Each new deal has amped up the excitement. But there is also anxiety that the field could be in a bubble. "It will be a cycle, and this cycle will eventually run its course," says Noubar Afeyan, managing partner at venture-capital firm Flagship Ventures in Cambridge, Massachusetts. "These things end up appearing and disappearing for reasons that people can only explain in hindsight."

Explanations abound for the cycle's upswing. For one thing, investor confidence has been bolstered by established biotech companies that have promising clinical data, says Michael Yee, an analyst at investment bank RBC Capital Markets in San Francisco, California. Yee says that some of these companies may be on the verge of launching blockbuster drugs that are projected to earn around \$10 billion a year in sales. Biotech firm Gilead of Foster City, California, for example, could soon receive approval for its treatments for chronic hepatitis C virus infections, and Bristol-Myers Squibb, based in New York, has created a stir with cancer treatments that target the immune system (see *Nature* 496, 14–15; 2013). The NASDAQ Biotechnology Index, comprising 124 biotech stocks, has risen 46% in the past 12 months.

Meanwhile, the US Food and Drug Administration (FDA), long seen as an obstacle in the industry, approved 39 new drugs in 2012 — the biggest annual tally in more than a decade. The FDA has also pledged to fast-track 'breakthrough' medicines for serious conditions. David Lubner, chief financial

BULLISH ON BIOTECH

Initial public offerings (IPOs) of stock have been completed by 24 biotech companies so far in 2013. With eight more planned this year, 2013 should break the record for proceeds set in 2000 — the last time a big boom turned to bust.



officer at the antibiotics company Tetrphase in Watertown, Massachusetts, says that the success of his firm's IPO in March might have been helped by legislation, passed last year, to ease the approval of drugs that fight antibiotic-resistant bacteria.

Universities also stand to benefit from the boom. When academic discoveries in the life sciences seed start-up companies, the universities license patents to the start-ups but charge little because young biotech companies rarely make a profit and are often cash-strapped. Instead, the universities receive company stock in the start-ups to supplement the small licensing fees. In a hot IPO market, the value of those equity stakes rises. "This market is a real boon for major research universities that have strong life-sciences programmes," says Donald Siegel, dean of the school of business at the State University of New York in Albany.

At the height of the technology boom in 2000, US universities made \$165 million by cashing in equity — money that is often split among individual inventors, their research departments and other university activities, such as technology-licensing offices. Last year, that tally was a respectable \$51 million, and Siegel says that this cashed-in equity represents less than 3% of universities' overall stake. Sean Flanigan, president of the Association of University Technology Managers in Deerfield, Illinois, says that universities stand to benefit in other ways from IPOs.

NATURE.COM
For more on biotech investment, see:
go.nature.com/ljj2bf

Companies can choose to funnel IPO money back into academia, by directly sponsoring research or licensing more technologies.

But the lucrative IPOs depend on continued enthusiasm — which could evaporate if investors decide that the market is too unstable. After a bubble bursts, investors can shun a field for years, says Chris Yung, an associate professor of finance at the University of Virginia in Charlottesville. "As a society, we seem to be pretty poor at determining when we've hit the right point to stop."

One sign of a bubble is when companies are valued more highly than their actual worth, says Yung. However, worth can be difficult to gauge for biotech companies that do not yet make a profit. For example, Agios, a pharmaceutical company in Cambridge, Massachusetts, currently has no drugs in clinical trials, yet the cancer-treatment firm raised \$106 million in its July IPO. Its stock shot up 74% on the first day of trading. Another example is Intrexon, a synthetic-biology company based in Germantown, Maryland. Yee says that Intrexon was one of the hottest biotech IPOs of the past two months, even though he is not entirely sure what the company aims to do scientifically and strategically.

But worth has many measures. Agios has a star-studded team of scientists at its helm and a \$150-million deal with Celgene, a large pharmaceutical company based in Summit, New Jersey. And Intrexon is run by Randal Kirk, a billionaire investor who has a history of reaping huge returns from biotech investments. Yee says that investors might feel good about supporting young companies that are targeting a well-defined population of patients, simplifying clinical trials and boosting the odds of success.

Clinical trial results are due from some of these companies in the next year — and investors will be keeping a close watch. "If the data comes out well, people are going to latch on and believe the dream is still alive," says Yee. "If it doesn't, investors will be reminded that a lot of this stuff is early stage — and it is risky." ■

SOURCE: RENAISSANCECAPITAL.COM

African genes tracked back

Method extends archaeological and linguistic data by tracing early human migration.

BY ERIKA CHECK HAYDEN

The first humans left Africa some 200,000 years ago, dispersing to populate the rest of the world. But this was not a one-way trip: some people came back. Scientists say that they have traced a reverse migration that, in two steps, carried genes from the rest of the world back to southern Africa, long before European colonizers arrived.

The findings are part of a flurry of research enabled by better tools to survey African genomes. For the first time, population geneticists can examine the complex history of human migration in Africa effectively, a field long dominated by the analysis of bones, artefacts and languages.

“Up until now this was mostly done based on linguistics and archaeology, and now we can use genetics to test ideas,” says Carina Schlebusch, a geneticist at Uppsala University in Sweden. “It’s a really exciting time for African genetics.”

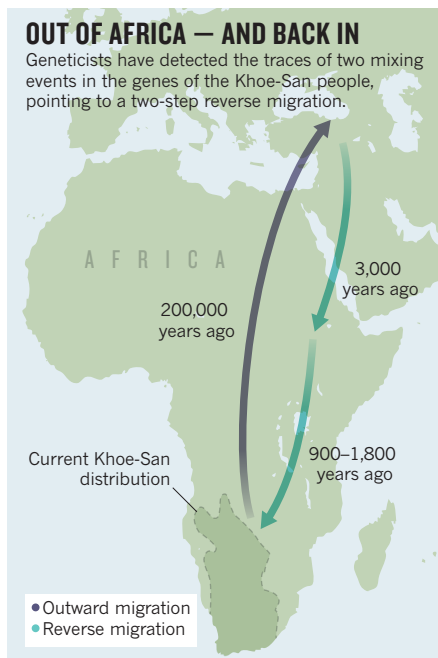
Genetic signs of reverse migrations had been spotted before. Over the past decade, evidence has accrued that people returned to Africa after the original ‘out of Africa’ event. And more recently, DNA data suggest that a small group of east Africans migrated to southern Africa to mix with the hunter-gatherers who lived there 2,000 years ago.

A new analysis connects these two migrations. In a paper posted on the preprint server arXiv on 30 July, population geneticists at Harvard Medical School in Boston, Massachusetts, report their use of a method to identify traces of this two-step migration within DNA of the minority Khoe-San ethnic groups of southern Africa (J. K. Pickrell *et al.* Preprint at <http://arxiv.org/abs/1307.8014>; 2013).

The researchers examined more than half a million sites of genetic variation in the genomes of around 1,000 people, including some 200 southern Africans representing 22 African groups. Turning to the Khoe-San, the team looked for regions of DNA in which individuals seemed to have inherited a chromosome from a different ethnic group at some time in the past.

These DNA regions contain blocks of genetic variants that are linked together in patterns similar to those seen in non-Khoe-San groups. The blocks are shorter and more scattered in the Khoe-San than in the presumed donor group, because the genetic recombination that occurs with each generation has, over time, fragmented the donated chromosome.

By measuring the extent of the fragmentation, the researchers could estimate how many



generations ago the chromosome had been introduced into Khoe-San genomes. They saw signs reflecting two waves of migration: one about 3,000 years ago, of non-Africans entering east Africa, and a second one 900–1,800 years ago, as east Africans migrated to southern Africa and brought non-African genes along with them (see ‘Out of Africa — and back in’).

Because of this two-step migration, some Khoe-San groups who were thought to have been genetically quite isolated actually carry 1–5% non-African DNA, report the study’s authors, led by David Reich and postdoctoral fellow Joseph Pickrell.

Geneticists had already been able to detect past instances of ‘admixture’, or interbreeding, but this was the first time that multiple admixtures had been detected within a single group’s genetic profile. “The most important finding is that this migration from east to south was a carrier for non-African genes,” says Luca Pagani, a geneticist at the Wellcome Trust Sanger Institute near Cambridge, UK, who has studied reverse migration to Ethiopia.

The finding makes sense in light of previous archaeological and linguistic studies, says Sarah Tishkoff, a genetic anthropologist at the University of Pennsylvania in Philadelphia. For instance, southern African speakers of

Khoe-Kwadi, a family of languages that traces its origins to east Africans who brought herding techniques to the south, had the highest proportion of non-African genes in the new study. Also, 2,000-year-old cattle and sheep bones and pottery characteristic of east African herding cultures have been found across southern Africa. “Perhaps this method could be applied to other cases where there isn’t good archaeological data,” says Tishkoff.

But the finding also raises questions about the identity of the eastern Africans who met the non-Africans, and of the non-Africans themselves. The non-African genes in the Khoe-San most closely resemble those found in modern southern Europeans. But the authors think it more likely that the carriers of those genes at the time actually crossed into Africa from the Middle East or the Arabian peninsula.

ADVANCE ON A CHIP

The research underscores the use of improved methods and richer data sets to explore the complex genetic history of Africa. Studies suggest that Africa is the most genetically diverse continent in the world and that the Khoe-San people have the world’s oldest genetic lineages. They are thought to be descendants of the first people to split from the other modern humans who began the journey out of Africa.

Despite the region’s outsized importance in human history, only in the past year has a significant amount of data on groups from all parts of Africa become available. Genomes of just a handful of the 2,000 African ethnic groups had previously been sequenced, partly because collecting DNA from remote African ethnic groups is an ethical and logistical challenge, and partly because most funding for genetic studies goes to medical, not anthropological, applications.

Compounding the problem, DNA microarrays designed to survey human genetic diversity were initially made to sample the variation seen in European and European-American genomes, and did not capture African diversity. That changed in 2011, when Reich and other geneticists worked with the company Affymetrix, based in Santa Clara, California, to design a gene chip — the Human Origins Array — that targets sites of variation seen in more diverse samples. Pickrell and Reich used the chip in their analysis, as have other surveys of African diversity.

That represents a welcome sea change, says Pagani. “Until now, we have been applying tools designed specifically for non-African people to African people.” ■

NATURE.COM
For more on the
out-of-Africa
migration, visit:
go.nature.com/scrdex



THE ORIGINS OF SPACE AND TIME

BY ZEEYA MERALI

Many researchers believe that physics will not be complete until it can explain not just the behaviour of space and time, but where these entities come from.

“Imagine waking up one day and realizing that you actually live inside a computer game,” says Mark Van Raamsdonk, describing what sounds like a pitch for a science-fiction film. But for Van Raamsdonk, a physicist at the University of British Columbia in Vancouver, Canada, this scenario is a way to think about reality. If it is true, he says, “everything around us — the whole three-dimensional physical world — is an illusion born from information encoded elsewhere, on a two-dimensional chip”. That would make our Universe, with its three spatial dimensions, a kind of hologram, projected from a substrate that exists only in lower dimensions.

This ‘holographic principle’ is strange even by the usual standards of theoretical physics. But Van Raamsdonk is one of a small band of researchers who think that the usual ideas are not yet strange enough. If nothing else, they say, neither of the two great pillars of modern

NATURE.COM

To see loop quantum gravity in action, visit.

go.nature.com/zusd5e

physics — general relativity, which describes gravity as a curvature of space and time, and quantum mechanics, which governs the atomic realm — gives any account for the existence of space and time. Neither does string theory, which describes elementary threads of energy.

Van Raamsdonk and his colleagues are convinced that physics will not be complete until it can explain how space and time emerge from something more fundamental — a project that will require concepts at least as audacious as holography. They argue that such a radical reconceptualization of reality is the only way to explain what happens when the infinitely dense ‘singularity’ at the core of a black hole distorts the fabric of space-time beyond all recognition, or how researchers can unify atomic-level quantum theory and planet-level general relativity — a project that has resisted theorists’ efforts for generations.

“All our experiences tell us we shouldn’t have two dramatically different conceptions of reality — there must be one huge overarching theory,” says Abhay Ashtekar, a physicist at Pennsylvania State University in University Park.

Finding that one huge theory is a daunting challenge. Here, *Nature* explores some promising lines of attack — as well as some of the emerging ideas about how to test these concepts (see ‘The fabric of reality’).

GRAVITY AS THERMODYNAMICS

One of the most obvious questions to ask is whether this endeavour is a fool’s errand. Where is the evidence that there actually is anything more fundamental than space and time?

A provocative hint comes from a series of startling discoveries made in the early 1970s, when it became clear that quantum mechanics and gravity were intimately intertwined with thermodynamics, the science of heat.

In 1974, most famously, Stephen Hawking of the University of Cambridge, UK, showed that quantum effects in the space around a black hole will cause it to spew out radiation as if it was hot. Other physicists quickly determined that this phenomenon was quite general. Even in completely empty space, they found, an astronaut undergoing acceleration would perceive that he or she was surrounded by a heat bath. The effect would be too small to be perceptible for any acceleration achievable by rockets, but it seemed to be fundamental. If quantum theory and general relativity are correct — and both have been abundantly corroborated by experiment — then the existence of Hawking radiation seemed inescapable.

A second key discovery was closely related. In standard thermodynamics, an object can radiate heat only by decreasing its entropy, a measure of the number of quantum states inside it. And so it is with black holes: even before Hawking’s 1974 paper, Jacob Bekenstein, now at the Hebrew University of Jerusalem, had shown that black holes possess entropy. But there was a difference. In most objects, the entropy is proportional to the number of atoms the object contains, and thus to its volume. But a black hole’s entropy turned out to be proportional to the surface area of its event horizon — the boundary out of which not even light can escape. It was as if that surface somehow encoded information about what was inside, just as a two-dimensional hologram encodes a three-dimensional image.

In 1995, Ted Jacobson, a physicist at the University of Maryland in College Park, combined these two findings, and postulated that every point in space lies on a tiny ‘black-hole horizon’ that also obeys the entropy–area relationship. From that, he found, the mathematics yielded Einstein’s equations of general relativity — but using only thermodynamic concepts, not the idea of bending space-time¹.

“This seemed to say something deep about the origins of gravity,” says Jacobson. In particular, the laws of thermodynamics are statistical in nature — a macroscopic average over the motions of myriad atoms and molecules — so his result suggested that gravity is also statistical, a macroscopic approximation to the unseen constituents of space and time.

In 2010, this idea was taken a step further by Erik Verlinde, a string theorist at the University of Amsterdam, who showed² that the

statistical thermodynamics of the space-time constituents — whatever they turned out to be — could automatically generate Newton’s law of gravitational attraction.

And in separate work, Thanu Padmanabhan, a cosmologist at the Inter-University Centre for Astronomy and Astrophysics in Pune, India, showed³ that Einstein’s equations can be rewritten in a form that makes them identical to the laws of thermodynamics — as can many alternative theories of gravity. Padmanabhan is currently extending the thermodynamic approach in an effort to explain the origin and magnitude of dark energy: a mysterious cosmic force that is accelerating the Universe’s expansion.

Testing such ideas empirically will be extremely difficult. In the same way that water looks perfectly smooth and fluid until it is observed on the scale of its molecules — a fraction of a nanometre — estimates suggest that space-time will look continuous all the way down to the Planck scale: roughly 10^{-35} metres, or some 20 orders of magnitude smaller than a proton.

But it may not be impossible. One often-mentioned way to test whether space-time is made of discrete constituents is to look for delays as high-energy photons travel to Earth from distant cosmic events such as supernovae and γ -ray bursts. In effect, the shortest-wavelength photons would sense the discreteness as a subtle bumpiness in the road they had to travel, which would slow them down ever so slightly. Giovanni Amelino-Camelia, a quantum-gravity researcher at the University of Rome, and his colleagues have found⁴ hints of just such delays in the photons from a γ -ray burst recorded in April. The results are not definitive, says Amelino-Camelia, but the group plans to expand its search to look at the travel times of high-energy neutrinos produced by cosmic events. He says that if theories cannot be tested, “then to me, they are not science. They are just religious beliefs, and they hold no interest for me.”

Other physicists are looking at laboratory tests. In 2012, for example, researchers from the University of Vienna and Imperial College London proposed⁵ a tabletop experiment in which a microscopic mirror would be moved around with lasers. They argued that Planck-scale granularities in space-time would produce detectable changes in the light reflected from the mirror (see *Nature* <http://doi.org/njfi>; 2012).

LOOP QUANTUM GRAVITY

Even if it is correct, the thermodynamic approach says nothing about what the fundamental constituents of space and time might be. If space-time is a fabric, so to speak, then what are its threads?

One possible answer is quite literal. The theory of loop quantum gravity, which has been under development since the mid-1980s by Ashtekar and others, describes the fabric of space-time as an evolving spider’s web of strands that carry information about the quantized areas and volumes of the regions they pass through⁶. The individual strands of the web must eventually join their ends to form loops — hence the theory’s name — but have nothing to do with the much better-known strings of string theory. The latter move around in space-time, whereas strands actually are space-time: the information they carry defines the shape of the space-time fabric in their vicinity.

Because the loops are quantum objects, however, they also define a minimum unit of area in much the same way that ordinary quantum mechanics defines a minimum ground-state energy for an electron in a hydrogen atom. This quantum of area is a patch roughly one Planck scale on a side. Try to insert an extra strand that carries less area, and it will simply disconnect from the rest of the web. It will not be able to link to anything else, and will effectively drop out of space-time.

One welcome consequence of a minimum area is that loop quantum gravity cannot squeeze an infinite amount of curvature onto an infinitesimal point. This means that it cannot produce the kind of singularities that cause Einstein’s equations of general relativity to break down at the instant of the Big Bang and at the centres of black holes.

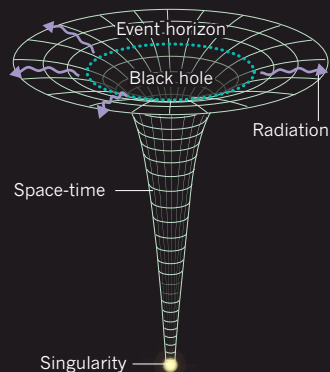
In 2006, Ashtekar and his colleagues reported⁷ a series of simulations that took advantage of that fact, using the loop quantum gravity

THE FABRIC OF REALITY

If space and time are not fundamental, then what is? Theoretical physicists are exploring several possible answers.

One clue

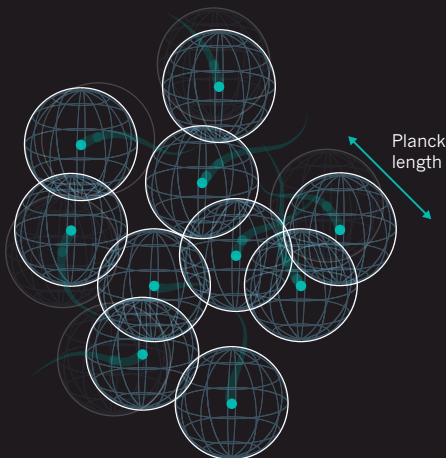
Quantum effects in the gravitational field of a black hole cause it to radiate energy as if it were hot, implying a deep connection between quantum theory, gravity and thermodynamics — the science of heat.



The black hole's mass is concentrated at a singularity of infinite curvature.

1. Gravity as thermodynamics

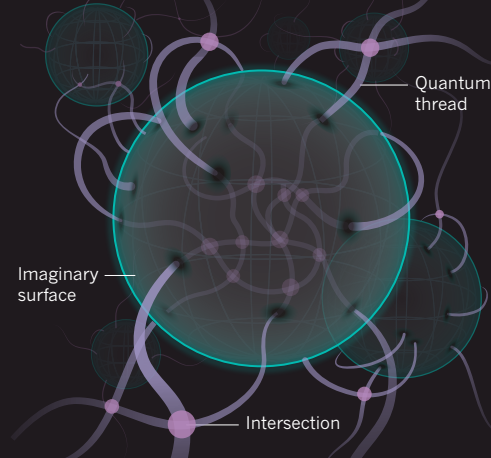
The equations of gravity can actually be derived from thermodynamics, without reference to space-time curvature.



This suggests that gravity on a macroscopic scale is just an average of the behaviour of some still-unknown 'atoms' of space-time.

2. Loop quantum gravity

The Universe is a network of intersecting quantum threads, each of which carries quantum information about the size and shape of nearby space.



Imagine drawing a closed surface anywhere in the network. Its volume is determined by the intersections it encloses; its area by the number of threads that pierce it.

version of Einstein's equations to run the clock backwards and visualize what happened before the Big Bang. The reversed cosmos contracted towards the Big Bang, as expected. But as it approached the fundamental size limit dictated by loop quantum gravity, a repulsive force kicked in and kept the singularity open, turning it into a tunnel to a cosmos that preceded our own.

This year, physicists Rodolfo Gambini at the Uruguayan University of the Republic in Montevideo and Jorge Pullin at Louisiana State University in Baton Rouge reported⁸ a similar simulation for a black hole. They found that an observer travelling deep into the heart of a black hole would encounter not a singularity, but a thin space-time tunnel leading to another part of space. "Getting rid of the singularity problem is a significant achievement," says Ashtekar, who is working with other researchers to identify signatures that would have been left by a bounce, rather than a bang, on the cosmic microwave background — the radiation left over from the Universe's massive expansion in its infant moments.

Loop quantum gravity is not a complete unified theory, because it does not include any other forces. Furthermore, physicists have yet to show how ordinary space-time would emerge from such a web of information. But Daniele Oriti, a physicist at the Max Planck Institute for Gravitational Physics in Golm, Germany, is hoping to find inspiration in the work of condensed-matter physicists, who have produced exotic phases of matter that undergo transitions described by quantum field theory. Oriti and his colleagues are searching for formulae to describe how the Universe might similarly change phase, transitioning from a set of discrete loops to a smooth and continuous space-time. "It is early days and our job is hard because we are fishes swimming in the fluid at the same time as trying to understand it," says Oriti.

CAUSAL SETS

Such frustrations have led some investigators to pursue a minimalist programme known as causal set theory. Pioneered by Rafael Sorkin, a physicist at the Perimeter Institute in Waterloo, Canada, the theory

postulates that the building blocks of space-time are simple mathematical points that are connected by links, with each link pointing from past to future. Such a link is a bare-bones representation of causality, meaning that an earlier point can affect a later one, but not vice versa. The resulting network is like a growing tree that gradually builds up into space-time. "You can think of space emerging from points in a similar way to temperature emerging from atoms," says Sorkin. "It doesn't make sense to ask, 'What's the temperature of a single atom?' You need a collection for the concept to have meaning."

In the late 1980s, Sorkin used this framework to estimate⁹ the number of points that the observable Universe should contain, and reasoned that they should give rise to a small intrinsic energy that causes the Universe to accelerate its expansion. A few years later, the discovery of dark energy confirmed his guess. "People often think that quantum gravity cannot make testable predictions, but here's a case where it did," says Joe Henson, a quantum-gravity researcher at Imperial College London. "If the value of dark energy had been larger, or zero, causal set theory would have been ruled out."

CAUSAL DYNAMICAL TRIANGULATIONS

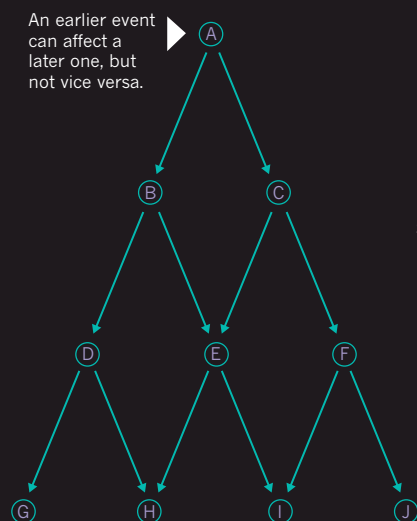
That hardly constituted proof, however, and causal set theory has offered few other predictions that could be tested. Some physicists have found it much more fruitful to use computer simulations. The idea, which dates back to the early 1990s, is to approximate the unknown fundamental constituents with tiny chunks of ordinary space-time caught up in a roiling sea of quantum fluctuations, and to follow how these chunks spontaneously glue themselves together into larger structures.

The earliest efforts were disappointing, says Renate Loll, a physicist now at Radboud University in Nijmegen, the Netherlands. The space-time building blocks were simple hyper-pyramids — four-dimensional counterparts to three-dimensional tetrahedrons — and the simulation's gluing rules allowed them to combine freely. The result was a series of bizarre 'universes' that had far too many dimensions

NIK SPENCER/NATURE

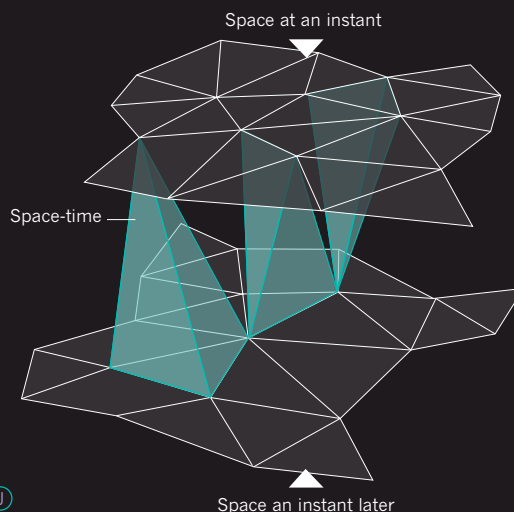
3. Causal sets

The building blocks of space-time are point-like 'events' that form an ever-expanding network linked by causality.



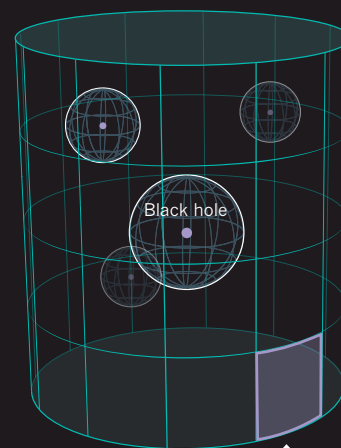
4. Causal dynamical triangulations

Computer simulations approximate the fundamental quantum reality as tiny polygonal shapes, which obey quantum rules as they spontaneously self-assemble into larger patches of space-time.



5. Holography

A three-dimensional (3D) universe contains black holes and strings governed solely by gravity, whereas its 2D boundary contains ordinary particles governed solely by standard quantum-field theory.



Anything happening in the 3D interior can be described as a process on the 2D boundary, and vice versa.

(or too few), and that folded back on themselves or broke into pieces. "It was a free-for-all that gave back nothing that resembles what we see around us," says Loll.

But, like Sorkin, Loll and her colleagues found that adding causality changed everything. After all, says Loll, the dimension of time is not quite like the three dimensions of space. "We cannot travel back and forth in time," she says. So the team changed its simulations to ensure that effects could not come before their cause — and found that the space-time chunks started consistently assembling themselves into smooth four-dimensional universes with properties similar to our own¹⁰.

Intriguingly, the simulations also hint that soon after the Big Bang, the Universe went through an infant phase with only two dimensions — one of space and one of time. This prediction has also been made independently by others attempting to derive equations of quantum gravity, and even some who suggest that the appearance of dark energy is a sign that our Universe is now growing a fourth spatial dimension. Others have shown that a two-dimensional phase in the early Universe would create patterns similar to those already seen in the cosmic microwave background.

HOLOGRAPHY

Meanwhile, Van Raamsdonk has proposed a very different idea about the emergence of space-time, based on the holographic principle. Inspired by the hologram-like way that black holes store all their entropy at the surface, this principle was first given an explicit mathematical form by Juan Maldacena, a string theorist at the Institute of Advanced Study in Princeton, New Jersey, who published¹¹ his influential model of a holographic universe in 1998. In that model, the three-dimensional interior of the universe contains strings and black holes governed only by gravity, whereas its two-dimensional boundary contains elementary particles and fields that obey ordinary quantum laws without gravity.

Hypothetical residents of the three-dimensional space would never

see this boundary, because it would be infinitely far away. But that does not affect the mathematics: anything happening in the three-dimensional universe can be described equally well by equations in the two-dimensional boundary, and vice versa.

In 2010, Van Raamsdonk studied what that means when quantum particles on the boundary are 'entangled' — meaning that measurements made on one inevitably affect the other¹². He discovered that if every particle entanglement between two separate regions of the boundary is steadily reduced to zero, so that the quantum links between the two disappear, the three-dimensional space responds by gradually dividing itself like a splitting cell, until the last, thin connection between the two halves snaps. Repeating that process will subdivide the three-dimensional space again and again, while the two-dimensional boundary stays connected. So, in effect, Van Raamsdonk concluded, the three-dimensional universe is being held together by quantum entanglement on the boundary — which means that in some sense, quantum entanglement and space-time are the same thing.

Or, as Maldacena puts it: "This suggests that quantum is the most fundamental, and space-time emerges from it." ■

Zeeya Merali is a freelance writer in London.

1. Jacobson, T. *Phys. Rev. Lett.* **75**, 1260–1263 (1995).
2. Verlinde, E. J. *High Energy Phys.* [http://dx.doi.org/10.1007/JHEP04\(2011\)029](http://dx.doi.org/10.1007/JHEP04(2011)029) (2011).
3. Padmanabhan, T. *Rep. Prog. Phys.* **73**, 046901 (2010).
4. Amelino-Camelia, G., Fiore, F., Guetta, D. & Puccetti, S. preprint at <http://arxiv.org/abs/1305.2626> (2013).
5. Pikovski, I., Vanner, M. R., Aspelmeyer, M., Kim, M. S. & Brukner, Č. *Nature Phys.* **8**, 393–397 (2012).
6. Ashtekar, A. preprint at <http://arxiv.org/abs/1201.4598> (2012).
7. Ashtekar, A., Pawłowski, T. & Singh, P. *Phys. Rev. Lett.* **96**, 141301 (2006).
8. Gambini, R. & Pullin, J. *Phys. Rev. Lett.* **110**, 211301 (2013).
9. Ahmed, M., Dodelson, S., Greene, P. B. & Sorkin, R. *Phys. Rev. D* **69**, 103523 (2004).
10. Ambjørn, J., Jurkiewicz, J. & Loll, R. *Phys. Rev. Lett.* **93**, 131301 (2004).
11. Maldacena, J. M. *Adv. Theor. Math. Phys.* **2**, 231–252 (1998).
12. Raamsdonk, M. V. *Gen. Rel. Grav.* **42**, 2323–2329 (2010).

COMMENT

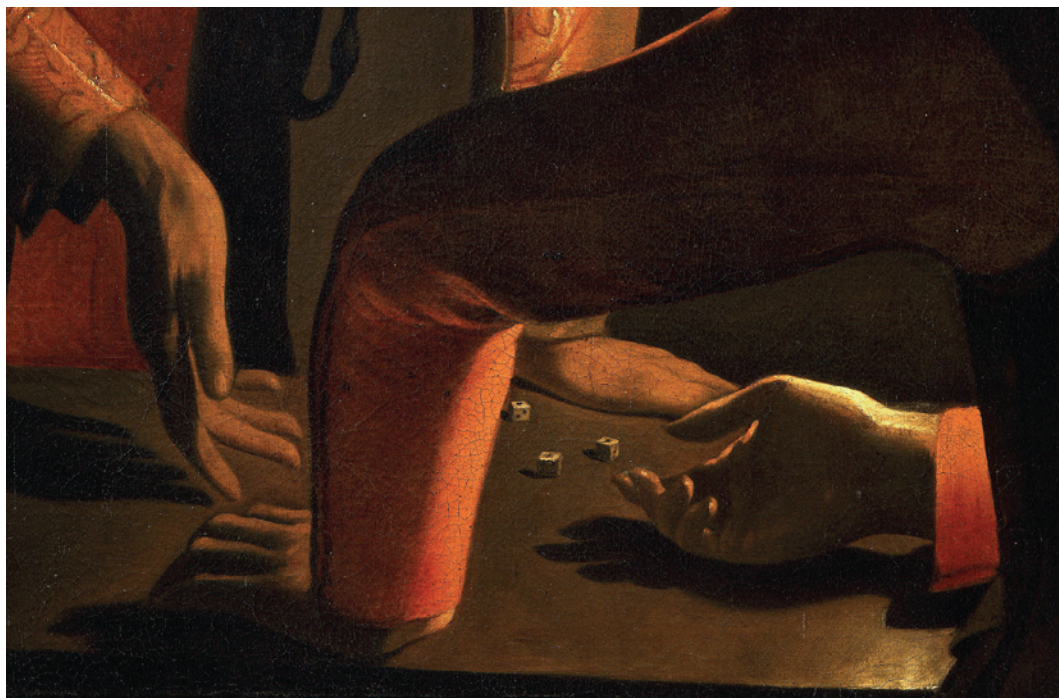
COMMUNICATION When metaphors confuse rather than clarify **p.523**

EXHIBITION A New York show of sound art challenges ideas about listening **p.526**

MUSIC A composer celebrates the geology of the Rocky Mountains **p.528**

OBITUARY Hugh Huxley, biophysicist who studied muscle contraction **p.530**

DETAIL FROM SAINT PETER'S DENIAL OF CHRIST BY GEORGES DE LA TOUR, 1650/
MUSÉE DES BEAUX ARTS NANTES/GIANNI DAGLI ORTI/THE ART ARCHIVE



Dice players in a detail from a 1650 painting by Georges de La Tour.

Value judgements

A mathematical paradox posed in a letter 300 years ago sowed the seed of economic theory by asking what money is worth, explains **George Szpiro**.

Three centuries ago, in September 1713, the Swiss mathematician Nikolaus Bernoulli wrote a letter to a fellow mathematician in France, the nobleman Pierre Rémond de Montmort. In it, Bernoulli described an innocent-sounding puzzle about a lottery. De Montmort found the problem so thought-provoking that within months, he published the letter in the second edition of his treatise *Essai d'analyse sur les jeux de hazard* ('Essay on the analysis of games of chance'). Little did the correspondents know that their exchange was the seed for the development of a fundamental concept of human decision-making, which would spawn the emergence of economics as a science.

The puzzle is about the gulf between what mathematicians expect from an uncertain event in the future, based on probability

theory, and what common sense tells us to do. The reverberations of the conundrum's eventual solution 25 years later are still felt today whenever a person chooses whether to buy home insurance, a bank manager decides what interest to charge a customer, or a financier ponders whether the likely returns on a risky venture warrant investing in it.

EXPECTATION MANAGEMENT

Bernoulli asked: if person A promises to give person B one coin if he throws six points on his first toss of a die, two coins if he gets six on the second throw, four coins if he gets six on the third, eight on the fourth and so on, then what can B expect to get? Bernoulli suggested that de Montmort would find "something very interesting" in this "easy" problem. Probability theory was then in its infancy.

The idea of 'expectation' had been coined 60 years earlier, in a correspondence between two French amateur mathematicians, the philosopher Blaise Pascal in Paris and the judge Pierre de Fermat in Toulouse. The pair had concluded that the expected value of an uncertain event is computed by multiplying the potential values with the probabilities of their occurrence.

De Montmort dismissed Bernoulli's problem, writing in November 1713 that it presented no difficulty whatsoever. All one had to do was to sum the relevant series. But de Montmort had missed the point. Reproachfully, Bernoulli retorted the following February that "you would have done well to seek the solution because it would have given you an occasion to make a very curious observation". ▶

► Reframing the problem from throwing a die to the simpler case of throwing a coin, Bernoulli proposed that the payout doubles each time the heads side of a coin does not appear. For example, Peter offers to pay Paul one gold ducat if a coin lands on heads on the first throw. If the first throw is tails, and the second is heads, Paul gets two ducats. If the first two throws land on tails, and heads appears on the third throw, Paul will get four ducats. If there are three tails in a row, and then heads, the payout will be eight ducats, and so on. What can Paul expect to win?

Following Pascal and Fermat's reasoning, the win is calculated as follows. The chance of the coin landing on heads on the first throw is $\frac{1}{2}$. The probability of the coin landing heads only on the second is $\frac{1}{4}$, the probability that heads will appear on the third throw, after two tails, is $\frac{1}{8}$, and so on. The expected win is the sum of the individual payouts (1, 2, 4, 8...) multiplied by the probabilities ($\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, $\frac{1}{16}$...).

The result is surprising. Each product — $1 \times \frac{1}{2}$, $2 \times \frac{1}{4}$, $4 \times \frac{1}{8}$, and so on — is a half. Because the series never ends, given that there is a real, if minute, chance of a very long run of tails before the first head is thrown, infinitely many halves must be summed. Shockingly, the expected win amounts to infinity. Incredulous, de Montmort wrote: "I am not able to believe that... the advantage to Paul be infinite." Here the matter rested for 14 years.

In May 1728, writing from London, the 23-year-old mathematician Gabriel Cramer from Geneva weighed in. "Mathematicians value money in proportion to its quantity, and men of good sense in proportion to the usage that they may make of it." This was a far-ranging insight. Adding a ducat to a millionaire's account will not make him happier, Cramer reasoned. On the assumption that any amount of money beyond 2^{24} (equal to 16,777,216) gives no extra utility to its owner, he summed Bernoulli's series up to that point and obtained a finite answer. An amount of 13 ducats, Cramer claimed, was the most that one should be willing to pay to participate in the game.

But Cramer soon recognized a flaw in his own argument. An extra ducat must have some utility, be it to a pauper or to a rich person. He found a fix. The usefulness of an extra coin is never zero, but simply less than that of the previous one — as wealth increases, so does utility, but at a decreasing rate. Assuming that utility increases with the square root of wealth, Cramer recalculated the expected win to be a little over 2.9 ducats.

Bernoulli replied with a subtly different solution. The reason for the infinity paradox, he said, was not that amounts beyond some large sum give little utility to a gambler, but that the gambler disregards minute probabilities. Bernoulli set probabilities smaller than $\frac{1}{32}$ to zero, and calculated the expected win as 2.5 ducats. Cramer's reply was contrite, even though his argument was just as legitimate.

At that point, Bernoulli turned to his younger cousin Daniel, a mathematician and physicist now known, among many other achievements, for Bernoulli's principle in fluid dynamics. Daniel's response has been lost, but we may surmise from Nikolaus's displeased answer that his cousin had sided with Cramer. The reason that Paul does not value any amount beyond 2^{24} ducats, Daniel apparently wrote, is that he fears that Peter, who must eventually pay the winnings, may not be rich enough to cover the nearly 17 million ducats if the first heads appeared only after 25 throws. Paul knows that he would never get more than that. Nikolaus dismissed his cousin's arguments, as he had Cramer's.

Frustrated, Daniel continued to work on the puzzle. Eventually, he came up with

their occurrence, he argued, it is the utility of each possible gain that must be multiplied by its probability. He suggested the logarithmic function as an indicator of the utility of wealth. The mean utility, converted back to its monetary value, is what a lottery is worth to the gambler. Again, Nikolaus dismissed these insights.

MEASURING RISK

Undaunted, Daniel sharpened his arguments over the next seven years. His 18-page paper on the measurement of risk was published in 1738 in *Commentaries of the Imperial Academy of Science of St Petersburg*. The solution to the 'St Petersburg paradox', as the conundrum is known, is still considered to be one of the seminal academic articles in economics. It was translated into German in 1896 and published in English in 1954 in the journal *Econometrica*.

Daniel encapsulated the probability scenario in a plot of utility versus monetary value, now known as a 'utility function' (see 'Risky business'). The curve rises — more money is always better than less — but does so at a decreasing rate, because an additional

amount provides less utility to a rich person than to a pauper. The curve's diminishing gradient implies that it is always worth paying a premium to avoid a risk.

The consequences of this simple graph are enormous. Risk aversion, as expressed in the concave shape of the utility function, tells us that people prefer to receive a smaller but certain amount of money, rather than facing a risky prospect. This, in turn, implies that homeowners are willing to pay a premium to insure their belongings, that investors expect higher returns for riskier assets and that borrowing rates are higher for a jobless person taking out a loan than for a professional.

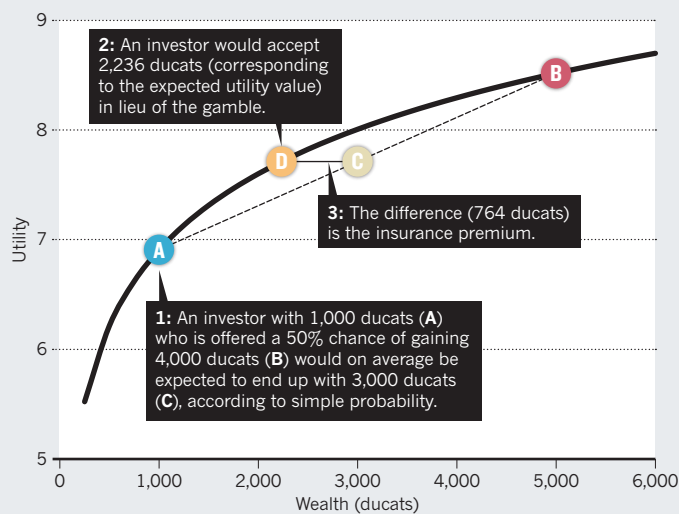
One of the first scholars to embrace Daniel's utility theory was the eminent French mathematician Pierre-Simon Laplace,

who included a lecture on 'expected utility' in his 1795 lecture course on probability theory at the École Normale Supérieure in Paris. The lecture was published in 1812 in his influential *Théorie analytique des probabilités* ('Analytic theory of probability').

In the mid-nineteenth century, utility theory received support from unexpected quarters. The physician Ernst Heinrich Weber and the psychologist Gustav Theodor Fechner found that people's physical sensations of, for example, weight, exhibit the same trait as their perception of wealth: the more weight one already carries, the heavier an extra load

RISKY BUSINESS

The concave shape of the utility function (here the natural logarithm) implies that every extra ducat grants less utility to its owner than the previous one.



another, more elegant, solution. In July 1731, he sent a manuscript to Nikolaus in which he explained that Fermat and Pascal's original proposition — calculating the expected value as the product of the outcomes and their probabilities — was the wrong approach. If this were so, he wrote, mathematical rules would govern people's decisions and everyone would agree on the correct choices.

Refining Cramer's insight, Daniel proposed that "the value of an item must not be based on its price, but rather on the utility it yields". Instead of multiplying the potential gains of a lottery with the probabilities of

must be in order for it to be noticed.

Among early economists, however, Daniel Bernoulli's theory was largely ignored until the twentieth century, when mathematician John von Neumann and economist Oskar Morgenstern — in their endeavour to lift economics from “plausibility considerations” to a mathematical science — provided an axiomatic framework for utility theory and decision-making in 1944. A few years later, economist Milton Friedman and statistician Leonard Savage, puzzled by the fact that many gamblers also take out household insurance, argued that utility functions have bulges and dents. Economist Harry Markowitz adapted utility functions in 1952 such that individuals consider their current wealth as a baseline, and are either risk-averse or risk-taking depending on whether potential losses or gains are relatively small, medium or large.

At about the same time, economist Maurice Allais pointed out that utility theory does not always account for people's behaviour. Faced with lopsided choices — for example, the certainty of \$1 million versus a chance of obtaining either hundreds of millions or nothing at all — people do not necessarily choose the ‘rational’ outcomes. Paradoxes such as these led sociologist and economist Herbert Simon to propose in the mid-1950s that humans are unable to gather all relevant information and to process it. As a result, they do not try to maximize their expected utility but, instead, set themselves more modest goals that will satisfy them.

In 1979, psychologists Daniel Kahneman and Amos Tversky developed prospect theory, which follows Daniel Bernoulli's lead but with some differences: losses hurt more than gains feel good; decisions depend on how the questions are framed; and probabilities are perceived to be smaller than they actually are, except for very small probabilities, which are perceived to be larger.

Three centuries on, Nikolaus Bernoulli's letter remains topical. Although he was not the one who provided the answer to the intriguing puzzle and, indeed, he resisted Cramer's and his cousin Daniel's explanations, it was his prompting of his friend to look deeply into the mathematics that set in motion a completely new way of thinking about risk, uncertainty and what money and wealth mean to people. ■

George Szpiro is a writer for the Swiss newspaper *Neue Zürcher Zeitung* and is based in New York City. He is currently writing a book about the history of decision-making.
e-mail: george.szpiro@nzz.ch



ILLUSTRATION BY HARRY CAMPBELL

Mind the metaphor

Imagery can help to bridge conceptual boundaries, but it can also cause trouble — as shown by the proliferation of engineering talk in biology, argues **Eleonore Pauwels**.

DNA barcodes, gene-shuffling, BioBrick parts and cells as hardware: synthetic biology is saturated with metaphors. And it is not an isolated case. In 1976, evolutionary biologist Richard Dawkins coined the term ‘selfish gene’ to explain a DNA-centred view of evolution. Ecologists built a whole metaphorical language around the idea of the ‘household of nature’, including terms such as competition and colonies. Beyond the natural sciences, the father of psychoanalysis, Sigmund Freud, described the restoration of an ego damaged by neurosis as the “reclamation of flooded lands”.

As a public-policy scholar, I have spent the past five years listening to synthetic biologists talk about their hopes, successes and failures. At first, I was intrigued by the pervasiveness of computing and engineering metaphors, both in conversations between scientists at the bench, and in policy discussions and public communications. Increasingly, I wanted to know what might be ‘lost

in translation’ between these metaphors and reality. In collaboration with my colleague Andrea Loettgers, a philosopher of science at the California Institute of Technology in Pasadena, I reviewed the use of metaphors in the laboratory and in the public sphere.

We looked at several sources, including more than 1,000 synthetic-biology articles, interviews with synthetic biologists and four years of US press coverage on the subject, as well as policy reports, US congressional hearings and bioethics-commission meetings. We found that although metaphors are essential in enabling science and in communicating research to the rest of the world, their use can also mislead the public, and even scientists themselves.

With the emergence of molecular biology in the 1940s, the idea of DNA as the ‘software of life’ became popular in the scientific community^{1,2}. Then, in the late 1990s, computer scientists, physicists and engineers were fuelled by the idea that they might ▶

► be able to direct cells in the same way that people program computers. In the laboratory, researchers started to use computing and engineering metaphors — switches, oscillators and logic gates, for instance — both to guide the design of synthetic constructs and to understand how natural systems function.

Almost immediately, scientists were confronted with the uncertainties and constraints of engineering in the cellular context. Engineering concepts and metaphors could serve only as inspiration; they were and are subject to much tinkering, owing to the complexity of biology. For instance, describing genetic systems as though they are electrical ones (whereby genes are switched on and off) works to a degree. But unlike switching on a light, which depends only on the flow of electricity, the activation of a particular gene depends on numerous parameters, and the precise effects of all of these different influences are often hard to pin down.

Despite the necessary fluidity surrounding their use, engineering metaphors have proved so robust as to create an identity among merging research communities. Indeed, the power of metaphors resides in their ability to serve as translational devices between different articulations of science — an essential function when cross-field collaboration results in the building of a new discipline, as has been the case for synthetic biology.

Scientists using metaphors among themselves are often aware of, and even careful to point out, the subtleties that could be misconstrued. Problems tend to arise when metaphors are used outside the laboratory.

Along with numerous journalists crowding the room at a May 2010 press conference in Washington DC, I listened transfixed as biologist Craig Venter announced that his team had become the first to build a self-replicating bacterial cell in the laboratory³ (see go.nature.com/xnq5h4). His words transformed a complex biological procedure into a science-fiction storyline: “This is the first self-replicating species we’ve had on the planet whose parent is a computer.”

Later that year, in a hearing convened by the US Committee on Energy and Commerce, Jay Keasling, a pioneer of synthetic biology based at the University of California, Berkeley, similarly described how synthetic biologists assemble “standardized well-characterized components from existing well-studied organisms, much like how one might assemble a computer from standard components such as a hard drive, sound card, motherboard and power supply”.

Faced with explaining the messy complexity and uncertainty of science to the public, it is understandable that scientists reach for metaphors. But discourse such as Venter’s and Keasling’s sends a message to policy-makers and laypeople that scientists can already

make biological systems that are reliable and controllable. It widens rather than closes the gap between scientific realities and the expectations of policy-makers and the public.

CONVEYING BELIEF

Psychologists at Stanford University in California showed in 2011 (ref. 4) that people’s views on how to manage crime varied drastically, depending on whether they were told that criminal activity is a ‘virus’ or a ‘wild beast’. It is because metaphors are so crucial to the perception of an idea that scientists need to use them with such care.

When new technologies emerge, optimism and enthusiasm often trump humility. In their excitement at making a discovery, many scientists, engineers and entrepreneurs suddenly believe that they can predict and control outcomes in complex physical and biological systems — and they frequently use metaphors to convey that belief.

One way to safeguard against runaway metaphors is to involve experts from diverse disciplines in the assessment of emerging technologies. Often, the sharing of expertise and outlooks helps to temper rhetoric and unpack what could get lost in translation. For instance, molecular biologist Bonnie Bassler of Princeton University in New Jersey argued in front of the 2010 US Presidential Commission for the Study of Bioethical Issues that the title of Venter and colleagues’ 2010 *Science* paper³ (‘Creation of a bacterial cell controlled by a chemically synthesized genome’) “does not represent the scientific findings” in it, and that “the authors did not create. They cloned.”

Organizers of the Science and Technology Innovation Program at the Woodrow Wilson International Center for Scholars in Washington DC have developed a strategy that could provide a model for practitioners across many disciplines. At the Wilson Center, experts and non-experts from different disciplines and sectors come together to discuss the science and implications of specific technological applications that are soon to be commercialized. Some of the latest examples include an arsenic biosensor and an algal biofuel.

Such collaborations between scientists, social scientists and policy-makers can drastically improve awareness of how powerful language can cut both ways. ■

Eleonore Pauwels is a researcher in the Science and Technology Innovation Program at the Woodrow Wilson International Center for Scholars, Washington DC, USA. e-mail: eleonore.pauwels@wilsoncenter.org

1. Kay, L. E. *Who Wrote the Book of Life?: A History of the Genetic Code* (Stanford Univ. Press, 1999).
2. Keller, E. F. *Making Sense of Life: Explaining Biological Developments with Models, Metaphors, and Machines* (Harvard Univ. Press, 2002).
3. Gibson, D. G. *Science* **329**, 52–56 (2010).
4. Thibodeau, P. H. & Boroditsky, L. *PLoS ONE* **6**, e16782 (2011).



Camille Norment's *Triplight* (2008), an exhibit in the *Soundings* show at New York's Museum of Modern Art.

ACOUSTICS

Subtleties of the sonic

Douglas Repetto explores the multilayered aesthetic of sound as art at New York's Museum of Modern Art.

Several years ago, a colleague noted similarities between the visualizations appearing on his computer screen when he played Beethoven, and graphs of the time-series data he was recording in his lab from cancer cells and sleeping human brains. Could he play back his data as sound, he asked? Yes: with some clever bit twiddling, you can turn any digital data source into sound, because a digital sound recording is simply a series of numbers. But is it music?

Entrenched controversies surround both the definition of music and the core relationships between sound, music and human perception. The bleeps and buzzes of bio-electrical recordings may sound nothing like Beethoven, but they would not be out of place on a dance-music track, and their structures can be similar to musical forms. Given the ambiguities, many artists working

Soundings: A Contemporary Score

Museum of Modern Art, New York City. Until 3 November 2013.

with sound avoid the term 'music' and call themselves sound artists, or simply artists. A wonderful collection of sound-oriented works by 16 mostly young artists is currently on display at New York's Museum of Modern Art. *Soundings: A Contemporary Score* is the first group show in the museum's history to focus on sound as artistic expression.

Soundings challenges the idea that listening is all about the ears. The rise of recording and playback technologies has left many thinking of sound (from music and audio books to rainforest-relaxation mixes) as best experienced in private through noise-cancelling headphones. But until about a hundred years ago, all sound was environmental: its source

was nearby and an individual's experience of it was driven by its social and environmental context. The works in the exhibition entirely dispense with the concept of private, ear-centric listening. Sergei Tcherepnin's *Motor-Matter Bench* (2013), for example, is rump-centric: a New York City subway bench has been wired up as a large speaker, with the heavy wooden seats filtering the sound. Sit on the bench and Tcherepnin's composition is transmitted through your bottom.

Nearby, Camille Norment's *Triplight* (2008), **pictured**, is completely silent: a flickering light inside the hollowed-out shell of an antique microphone casts a beguiling ribcage shadow on the walls, evocative of a smoky jazz club. That such a simple visual gesture can evoke a complex, multi-modal scene is not surprising, given the tangled connections between our senses,

CAMILLE NORMENT AND GASSER GRUNERT GALLERY, NEW YORK

perceptions and memories.

Some artists work directly with sound as a physical material, finding ways to alternately elucidate and confound our physical and perceptual understanding of sonic phenomena. In *wellenwanne lfo* (2012), Carsten Nicolai plays inaudible, low-frequency sounds through four small transducers resting in a shallow trough of water. The rippled interference patterns are reflected by a mirror and presented in an upright, frame-like light box — inaudible sound as a gently undulating abstract photograph.

Tristan Perich's monumental *Microtonal Wall* (2011) is certainly audible, and what you hear is both exhilarating and deeply confusing. A grid of 1,500 small speakers is mounted in an aluminium frame along a hallway. Each speaker plays a fixed frequency, arranged low to high from the bottom-left to top-right of the grid. The tones span four octaves, and, as expected, the effect of such dense sound is almost white noise. But if you get an ear close enough (and everyone does), you suddenly discover a world of rich physical and psychoacoustic effects: endless microglissandi, wild acoustic beating and bewildering spatial movements. Our wet brains and the messy physical world quickly turn a clean, precise stimulus into a riot of perception.

Christine Sun Kim's soundless works on paper play with the translation of sound and language concepts into visual marks. In *Scores and Transcripts* (2012), she presents a set of playful but poignant drawn interpretations of American Sign Language phrases and musical dynamics. From a distance, the red swoop of *All. Night.* (2012) reads as a giant cartoon chin. *Pianoiss ... issmo (Worse Finish)* (2012) is a classic bifurcation diagram rendered with tiny pianissimo marks. Deaf since birth, Kim's seemingly offhand drawing style belies a rich understanding of the mostly unconscious ways in which sound mediates physical and social relationships in the hearing world.

Jana Winderen takes a literal approach to translating sounds between worlds in her 16-speaker installation *Ultrafield* (2013). She recorded the ultrasonic clicks, whistles and whirrs of insects, fish and bats, and then digitally shifted their frequencies down into the range of human hearing before mixing them into a densely layered, half-familiar, half-exotic sound world. It's the sound of a science-fiction superhighway running through a shrieking, burbling rainforest. Or maybe the sound of a human brain, struggling to make sense of itself as a species that enjoys shrilling insects and string trios in equal measure. ■

Douglas Repetto is an artist, teacher and director of the Sound Art MFA programme at Columbia University in New York.
e-mail: dr501@columbia.edu

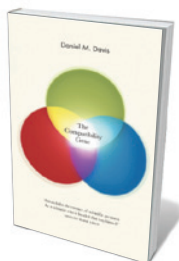
Books in brief



Paul Lauterbur and the Invention of MRI

M. Joan Dawson MIT PRESS (2013)

As light-bulb moments go, Paul Lauterbur's was one of the quirkiest. On 2 September 1971, the chemist was eating a hamburger when he realized that images could be derived from nuclear magnetic resonance signals, then used to detect tumours in tissue samples. The discovery paved the way for magnetic resonance imaging, which allows diagnosis of a multitude of conditions without the need for ionizing radiation. Molecular biologist Joan Dawson, Lauterbur's widow, deftly interweaves biography and breakthroughs with the findings of others in the field, such as Raymond Damadian.



The Compatibility Gene

Daniel M. Davis ALLEN LANE (2013)

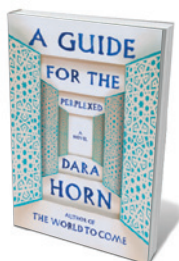
Eminent immunologist Daniel Davis tells the story of the "molecular mark that distinguishes each of us as individuals". The genes of the major histocompatibility complex (MHC) are big players in the immune system, and are implicated in conditions from ankylosing spondylitis to multiple sclerosis. Starting with the work of Peter Medawar and Peter Gorer on immunology, transplantation and the MHC, Davis ranges energetically through the research. Cultural references and anecdotes abound, but the book's shifts in tone between breezy and technical may unsettle some.



Their Fate is Our Fate: How Birds Foretell Threats to Our Health and Our World

Peter Doherty THE EXPERIMENT (2013)

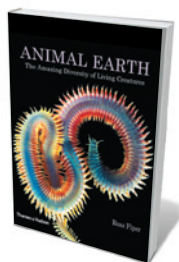
Like billions of coal-mine canaries, avian species tell us much about environmental crises, Nobel-prizewinning immunologist Peter Doherty reminds us. He explores how birds reflect the progress of epidemics and environmental change, framed by his own adventures in disease ecology. From 'sentinel chickens' — used to monitor the spread of arbovirus — to the migratory red knot *Calidris canutus rufa*, whose decline points to trouble down the food chain, Doherty flies with us through key vistas of science.



A Guide for the Perplexed: A Novel

Dara Horn W. W. NORTON (2013)

Computer science and medieval philosophy mesh in Dara Horn's accomplished novel about digital dangers and the nature of memory. Software supremo Josie Ashkenazi's program Genizah (Hebrew for 'repository of sacred texts') creates palaces of memory housing images and documents. In post-Arab Spring Cairo, she is brutally kidnapped. That headlong narrative is woven through with the real-life stories of scholar Solomon Schechter, who discovered a famous genizah in Cairo, and twelfth-century Jewish philosopher Moses Maimonides, whose treatise gives the novel its name.



Animal Earth: The Amazing Diversity of Living Creatures

Ross Piper THAMES AND HUDSON (2013)

In a seabed near you may lurk *Tubiluchus priapulids* — penis worms resembling shower heads with tails. Zoologist Ross Piper presents a parade of charismatic yet overlooked fauna in this sumptuous volume. The bioluminescent comb jelly *Beroë abyssicola* — not unlike a glowing papaya — and the candy-coloured twists of annelid worm *Alitta virens* are astounding enough. But the Persian carpet flatworm *Pseudobiceros bedfordi* is an aesthete's delight, sporting a black-and-scarlet pattern straight out of haute couture. [Barbara Kiser](#)



Q&A Jeffrey Nytch

Composer of deep time

For the 125th anniversary of the Geological Society of America, the Colorado-based musician Jeffrey Nytch has composed a symphony celebrating the geology of the American Rocky Mountains. He talks about Formations, which will be premiered by the Boulder Philharmonic Orchestra on 7 September.

How did you become interested in geology?

I actually studied geology and music. I wrote a piece at university that was directly inspired by geology. I looked at mass-spectrometry profiles of different minerals and created an arbitrary formula to transfer their resonance frequencies into the audible spectrum. I don't think it was a very good piece, but it was my first foray into trying to join the two things together. For me, art and science are both trying to do the same thing: understand the world and our place in it. Whether we're talking about that in a metaphorical or a concrete sense, I think they're both after the same basic thing.

How did the commission for *Formations* arise?

This was my crazy idea. I went to a community seminar about the geology of the Rocky Mountains, and I heard that the Geological Society of America [GSA], based in Boulder,

Colorado, was having its 125th anniversary this year. The light just went on for me. I went to the Boulder Philharmonic and said, "You should commission a symphony from me, inspired by the geology of the West". And the GSA funded it. They had been looking for some keystone event to tie their celebration together, and for them this was totally out of the box.

How did you shape *Formations*?

The orchestra wanted it to be about 25 minutes long, and it immediately became clear that I couldn't capture the whole geological story. I had to step back and say, "What are some key episodes that would make good music?" I began at the beginning, with the early Precambrian eon [between 4,600 million and 542 million years ago]. And the end needed to be something pretty grand, so I picked the orogenies [mountain-forming episodes] that built the modern Rockies. I

then had the two bookends, but what comes in the middle is sometimes the trickiest part. I wanted to weave human stories into it, and that was when I decided that the two that relate directly to geology are the gold and silver rushes of the nineteenth century, and oil and gas drilling in the twentieth century. That's when the two middle movements began to take shape.

Are there common musical themes that run through the movements?

There is a motif associated with uplift, and a motif associated with Precambrian rock that keeps transforming in many different ways as it is eroded again and again. I wanted to try to create musical narratives that were more about how the rocks formed.

How do you convey geological deep time through music?

The bad news is that the time spans we're talking about are so vast that we can't even begin to wrap our brains around them. The good news is that music is a kind of metaphor for that, because unlike a painting, it unfolds in time. I try to use long musical gestures to reflect long expanses of time. I'm trying to guide the listener towards a broad sense of how things formed. When you look at a geological formation, you're seeing it as it exists now — as a snapshot in time. But you also see all the time that was required to form it, the uplift and erosion and everything else — time compressed into this single instant. And so I think that's why I can get away with it in music: I can tell the story in 25 minutes.

What were the hardest parts to put to music?

Volcanoes are hard. I had to try to create the energy behind a volcanic eruption, as opposed to depicting the spewing of something. But the hardest aspects to portray were the gold and silver rushes. In the end I resorted to some tricks. In one section I wanted to depict the miners panning for gold, so I created a miner's pan with rocks that rattle around in it. And in hydrothermal intrusions, water comes bubbling up to the Earth's surface, so I'll have all of the brass section blow through their instruments to make a whooshing sound.

What do you hope people who don't live near the Rockies might get from your work?

The story of our evolving planet is a universal one, and I hope it will inspire people to think about how that story fits in with where they live. Ultimately, this is about trying to give Earth a voice. Earth has told us the story of how it formed: it's in the rocks. *Formations* is about trying to interpret that in a language we can understand. ■

INTERVIEW BY ALEXANDRA WITZE

Correspondence

Arctic: speed of methane release

Gail Whiteman and colleagues suggest that the opening up of the Arctic Ocean could bring more economic costs than benefits, owing to climatic effects resulting from a sudden release of 50 gigatonnes of methane from the area (*Nature* **499**, 401–403; 2013). However, our literature review of the impact of sea-ice decline on Arctic greenhouse-gas exchange indicates that methane release is likely to be more gradual because of a slow rate of heat penetration into the subsea permafrost (see F. J. W. Parmentier *et al.* *Nature Clim. Change* **3**, 195–202; 2013). We therefore believe that the proposed scenario is unlikely.

Although the Arctic Ocean represents a substantial source of methane, there are still many unknowns. Any research that assumes a large increase in emissions from that region should therefore include ample discussion of the uncertainties relating to this source.
Frans-Jan W. Parmentier, Torben R. Christensen *Lund University, Sweden.*
frans-jan.parmetier@nateko.lu.se

Arctic: uncertainties in methane link

We disagree with Gail Whiteman and colleagues that there is “likely” to be a large and sudden release of methane from the East Siberian Arctic Shelf (*Nature* **499**, 401–403; 2013).

Such an event would require an almost 1,000-fold regional increase in methane emissions from thawing permafrost, which would be inconsistent with geological evidence: although the concentration of atmospheric methane rose in response to abrupt warming during recent deglaciations, isotopic methane measurements do not indicate that the gas came from marine gas hydrates during these periods (H. Fischer *et al.* *Nature* **452**,

864–867; 2008).

Sea-floor temperatures determine the stability of methane hydrate reservoirs. To our knowledge, there is no evidence that these temperatures have changed significantly in the past few decades, or that a sudden change is imminent.

Although elevated methane concentrations have been observed in water and the atmosphere above the East Siberian Arctic Shelf, it is not clear whether these have been caused by recent warming or by natural processes linked to glacial–interglacial changes (V. V. Petrenko *et al.* *Science* **329**, 1146–1147; 2010). The global increase in methane recorded in the past few years does not seem to have been caused by enhanced Arctic emissions.

We welcome studies that quantify the economic impact of global warming, but they need to be accompanied by a realistic assessment of the uncertainties.
Dirk Notz, Victor Brovkin *Max Planck Institute for Meteorology, Hamburg, Germany.*
dirk.notz@zmaw.de
Martin Heimann *Max Planck Institute for Biogeochemistry, Jena, Germany.*

Gail Whiteman, Chris Hope and Peter Wadhams respond: Since 2005, accelerating sea-ice retreat in the Arctic has exposed the sea bed to much warmer conditions than those considered in earlier studies that anticipated a slow release of methane. Summer water temperatures off Siberia now climb to several degrees above 0°C (see, for example, N. R. Bates *et al.* *Biogeosciences* **10**, 5281–5309; 2013), causing the upper layers of offshore permafrost to melt more rapidly than they did a decade ago.

We have rerun our model with the same total quantity of emitted methane, but released over 50 or 75 years rather than 10 years. The results show no reduction in the total cost to society — in fact, the discounted costs over time would be larger.

Arctic: warming impact is uneven

We support the call for more rigorous integrated modelling of the costs and benefits of Arctic warming (G. Whiteman *et al.* *Nature* **499**, 401–403; 2013). Such models should examine the distribution of these costs and benefits within, as well as between, countries.

Melting permafrost, as Whiteman *et al.* show, gives rise to costly methane emissions. In addition, it damages the infrastructure needed for shipping and transportation, and for gas, oil and mineral mining, at high northern latitudes; this is because previously solid ground loses its structural integrity.

The Madrid-based humanitarian organization DARA expects the cost of this damage in the Arctic to reach US\$80 billion annually by 2030 (see go.nature.com/vnlzax). These losses will be unevenly distributed, with more than 90% likely to occur in Russia. The benefits from natural-resource extraction will also vary significantly between countries.

Indigenous peoples in the far north, who are already marginalized, will be subjected to further economic hardship and deteriorating conditions for their traditional hunting and land-use practices.

We believe that the uneven distribution of costs and benefits of Arctic warming calls for a new approach to governance in the region. This should ensure equitable compensation and support for those affected most.

Janis Hoberg, Francisco Ascuí *University of Edinburgh, UK.*
janis.hoberg@gmail.com

Brain stimulation has a long history

Self-administered transcranial electrical stimulation (TES) is not new (see *Nature* **498**, 271–272; 2013). Between the 1740s and 1930s, there was widespread

and unregulated commercial availability of TES machines for private and domestic use (see, for example, I. R. Morus in *Frankenstein's Children*, Princeton University Press; 1998).

During Victorian and Edwardian times, TES machines that dispensed static, frictional, faradic or battery electrical current could be bought everywhere. Some physicians, therapists and patients claimed that TES could generate feelings of euphoria and even improve mental performance.

Physicians of the period recommended that currents of no more than 5–10 milliamps should be applied to the head because of the risks of burning and shock, although many users chose to experiment. Side effects included headaches, flashes of light, dizziness and nausea, especially when connections were imperfect or broken.

The consequences could be even more serious. The Dutch physician Jan Ingenhousz knocked himself unconscious when performing electrical experiments in 1783, and Benjamin Franklin suffered retrograde amnesia after accidentally administering an electric shock to his head.

As today, some medical practitioners of the time believed the benefits of TES to be exaggerated.

Roi Cohen Kadosh *Department of Experimental Psychology, University of Oxford, UK.*

Paul Elliott *School of Humanities, University of Derby, UK.*
p.elliott@derby.ac.uk

CONTRIBUTIONS

Correspondence may be submitted to correspondence@nature.com after consulting the guidelines at <http://go.nature.com/cmchno>. Alternatively, readers may comment online: www.nature.com/nature.

Hugh Huxley

(1924–2013)

Biophysicist who established the mechanism of muscle contraction.

In a career spanning more than 65 years, Hugh Huxley achieved his lifelong ambition of understanding how muscles contract.

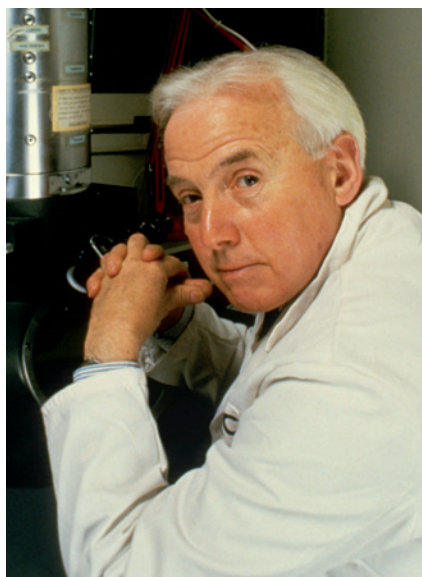
Huxley, who died on 25 July after a heart attack, was born in 1924 in Cheshire, UK, and studied physics at Christ's College at the University of Cambridge, UK. His education was interrupted by service in the Royal Air Force from 1943 to 1947, testing height-to-surface (H2S) radar — a ground-scanning system used by bomber aircraft. There, he learned the importance of doing work himself, and experimenting with electrical and mechanical devices became a lifetime passion. Huxley was able to fix a wildly inaccurate version of H2S developed for east Asia by switching from a voltage-controlled display system, in which overheating was a problem, to a current-controlled system. He described the feat to me some 60 years later as his first 'eureka moment' — for which he was made a Member of the Order of the British Empire (MBE) in 1948.

With the confidence of youth, that same year, Huxley chose to study muscle structure using X-ray methods for his PhD. He joined Max Perutz's group at the Medical Research Council (MRC) Unit for Research on the Molecular Structure of Biological Systems in Cambridge, with John Kendrew as his supervisor. As with radar, X-ray crystallography was in its infancy, and this project would test Huxley's talent for experimentation, attention to detail and determination to succeed.

Little was known then about the detailed structure of muscle. A pattern of dark and light bands had been identified with light microscopy, and the proteins actin and myosin were known to form filaments that interacted, although their exact location in muscle cells was unclear. From the analysis of X-ray patterns, Huxley identified big changes in the intensities of some reflections when muscles were contracted (when ATP, which provides cells with energy, is removed). He deduced the presence of a hexagonal array of two different types of filaments, and concluded that muscle contraction involved physical links between myosin and actin.

Needing visual evidence, he went to the Massachusetts Institute of Technology in Cambridge in late 1952. He was joined by the biophysicist Jean Hanson and, using both electron and light microscopy, they demonstrated the presence of two sets of overlapping filaments and found that the

salt solutions that extract myosin removed only the optically dense A-bands (another eureka moment). Their 1953 *Nature* paper, together with earlier X-ray work, provided the key evidence for the 'sliding filament' model of muscle contraction.



Muscle experts were sceptical, but a chance meeting with fellow Brit Andrew F. Huxley (no relation) at Woods Hole in Massachusetts lent Hugh's theory strong support. Andrew Huxley had come to similar conclusions using interference microscopy to analyse single muscle fibres, although without the fine detail revealed by electron microscopy. The two Huxleys agreed to publish simultaneously, and the papers appeared in *Nature* in May 1954, with Hugh Huxley's the more substantial, in my view. This was the defining moment in the study of muscle.

What made the filaments slide? In 1958, Huxley discovered myosin 'cross-bridges'. These attach to actin in a herringbone fashion in the absence of ATP, and detach when ATP is present. They drive muscle motion like rowing oars.

In 1962, Huxley returned to Britain to join the newly opened MRC Laboratory of Molecular Biology in Cambridge (LMB), where he later became deputy director. Over the next 50 years he was relentless in his quest to understand cross-bridge movement. His seminal 1969 *Science* paper described the swinging cross-bridge model, and subsequent computer reconstructions showed

how calcium ions regulated cross-bridge interaction. In 1970, using intense X-ray synchrotron radiation, Huxley began time-resolved studies of cross-bridge movement.

Huxley's move in 1988 to Brandeis University in Waltham, Massachusetts, as director of the Rosenstiel Basic Medical Sciences Research Center, gave him access to even more powerful X-ray sources, particularly at Argonne National Laboratory in Illinois. Improved detectors with high-sensitivity charge-coupled devices (used in digital cameras) enabled him to collect in milliseconds data that would have taken hours when he started out. With the atomic structures of actin and the myosin cross-bridge, along with other evidence, he wrote in the *European Journal of Biochemistry* in 2004 that he finally had direct evidence for the type of cross-bridge movement that he had postulated nearly 50 years earlier. It gave him great satisfaction that the mechanism by which myosin functions in muscle is replicated in almost all biological movement. His love for experimentation had delivered beyond his wildest dreams.

Hugh admired independence of spirit and enterprise, and he led by example. His personality reflected his strong humanist views. Outside science, he enjoyed travelling, sailing and skiing; he shared with his wife Frances a passion for theatre, relished a fine burgundy and loved the operas of Giacomo Puccini.

For the LMB's 50th anniversary in 2012, and with a new and greatly enlarged LMB opened in 2013, Hugh asked more than 40 academics who had visited the lab between 1957 and 1988 to write about how it had influenced their careers. These compiled writings, *Memories and Consequences* (MRC, 2013), provide fascinating reflections. The cell biologist Richard McIntosh described the atmosphere at the LMB as exuding "quality of thought, care in execution and energy in pursuit of excellence, reflecting the power of intelligent collaboration between labs with different kinds of expertise". There could be no more fitting epitaph for Hugh Huxley. ■

Alan Weeds was hired by Hugh Huxley in 1967 to establish a biochemistry group to study myosin, and worked on motile systems at the Laboratory of Molecular Biology in Cambridge, UK, for more than 40 years. He retired in 2005 and is a life fellow of Trinity College, University of Cambridge, UK. e-mail: agw22@cam.ac.uk

PETE ADDIS/PHOTOFUSION

FORUM: BIOGEOCHEMISTRY

Ancient algae crossed a threshold

The finding that the shells of certain algae can contain a signature of low levels of atmospheric carbon dioxide has prompted the discovery of the emergence of this signature in the fossil record. Here, experts discuss the implications of this for climate science and ocean ecology. [SEE LETTER P.558](#)

THE PAPER IN BRIEF

- Coccolithophores are marine algae that use inorganic carbon for photosynthesis and for calcification — the precipitation of calcium carbonate to produce an exoskeleton made up of plates called coccoliths (Fig. 1).
- In this issue, Bolton and Stoll¹ report that coccolithophores allocate more inorganic carbon, in the form of bicarbonate from sea water, to photosynthesis than to calcification

when atmospheric levels of carbon dioxide are low.

- This change in allocation causes a difference between the carbon-isotope signature of small and large coccoliths.
- The authors detect such a difference in the fossil record beginning about 7 million years ago.
- They conclude that a global decrease in the concentration of atmospheric CO₂ must have occurred at that time.

Climate lessons

RICHARD D. PANCOST
& MARCUS P. S. BADGER

Bolton and Stoll have developed a tool to reconstruct ancient carbon dioxide concentrations on the basis of differences in the carbon isotopic composition of large and small coccoliths. Their method is not a direct CO₂ barometer, but is based on what seems to be a threshold response of coccolithophores, which undergo a change in physiology as CO₂ levels increase above about 375–575 parts per million (p.p.m.). The authors used their approach to test directly whether CO₂ concentration has slowly decreased from higher levels 10 million to 12 million years (Myr) ago. With CO₂ levels currently on the rise, the answer has implications for future climate. The authors' work goes beyond being just another (much needed) approach to reconstructing CO₂ concentrations, and has implications for our understanding of how phytoplankton adapt to the world around them.

Ancient CO₂ levels can be determined from bubbles in ice cores, but these records extend back by only about 1 Myr. For the rest of Earth's history, proxy approaches are necessary². Such proxies can be based on the stomatal density of fossil leaves, the carbon-isotope discrimination of photosynthesis as recorded by alkenones (compounds produced exclusively by some marine coccolithophores) or the boron isotopic composition of marine

plankton. Ideally, multiple methods are used to constrain uncertainties.

Unfortunately, the proxy data from 12 Myr to 5 Myr ago are largely limited to the alkenone approach³, and even those data are restricted, originating largely from a single site in the southwest Pacific Ocean. That record suggests that, during this period, CO₂ levels were rather low (below 300 p.p.m.) and relatively stable, or increasing slightly. Combining that observation with alkenone data from other sites and from more recent time periods yields a record suggesting that CO₂ levels peaked at around 4–5 Myr ago⁴. By contrast, sea surface temperatures steadily dropped over the same period and continental ice sheets expanded, causing some researchers⁵ to argue that climate was decoupled from CO₂ levels.

Using their approach, Bolton and Stoll have produced data that present a fundamentally different model for CO₂ evolution over the past 12 Myr from that derived from the combined alkenone record. Their data indicate that CO₂ levels were elevated 12–7 Myr ago and decreased 7–5 Myr ago. Crucially, this suggests that CO₂ levels were indeed coupled to ocean temperature during much of the past 12 million years.

One implication of the authors' work is for the aforementioned alkenone CO₂ barometer, because coccoliths and alkenones derive largely from the same organisms. The alkenone proxy is based on a theoretical framework⁶ that does not include active uptake of CO₂ by coccolithophores, but Bolton and Stoll's work indicates that this framework is flawed. Although this

does not necessarily mean that the empirical relationship on which the alkenone proxy is based is incorrect, the way in which the proxy is interpreted and extrapolated to ancient settings could be overly simplistic. This might help to explain the apparent disconnect between climate- and alkenone-based CO₂ proxies throughout the Miocene epoch (about 23–5 Myr ago).

It is useful to place the authors' conclusions in the context of the ongoing, human-caused increase in CO₂, which has risen from about 280 p.p.m. in pre-industrial times to almost 400 p.p.m. at some sites this summer⁷, probably for the first time in millions of years⁴. Bolton and Stoll show that, during the past 12 Myr, periods of higher CO₂ levels were almost always characterized by markedly higher temperatures and smaller ice sheets than those of today.

Richard D. Pancost and Marcus P. S. Badger are in the Organic Geochemistry Unit, The Cabot Institute, School of Chemistry, University of Bristol, Bristol BS8 1TS, UK. e-mails: r.d.pancost@bristol.ac.uk; marcus.badger@bristol.ac.uk

Sea changes

JOHN REINFELDER

It was recently reported⁸ that at low concentrations of CO₂ coccolithophores seem to divert 'pumped' bicarbonate — sea water bicarbonate that has been actively transported into cells — away from calcification and towards photosynthetic production of organic carbon. Bolton and Stoll provide independent confirmation of this shift in carbon metabolism for a variety of coccolithophore species from the modern ocean and from the geological past, and they show that the expression of this shift depends on cell size. So what does this tell us about ocean ecology?

Most studies of the regulation of photosynthesis and calcification by CO₂ in coccolithophores have used various strains

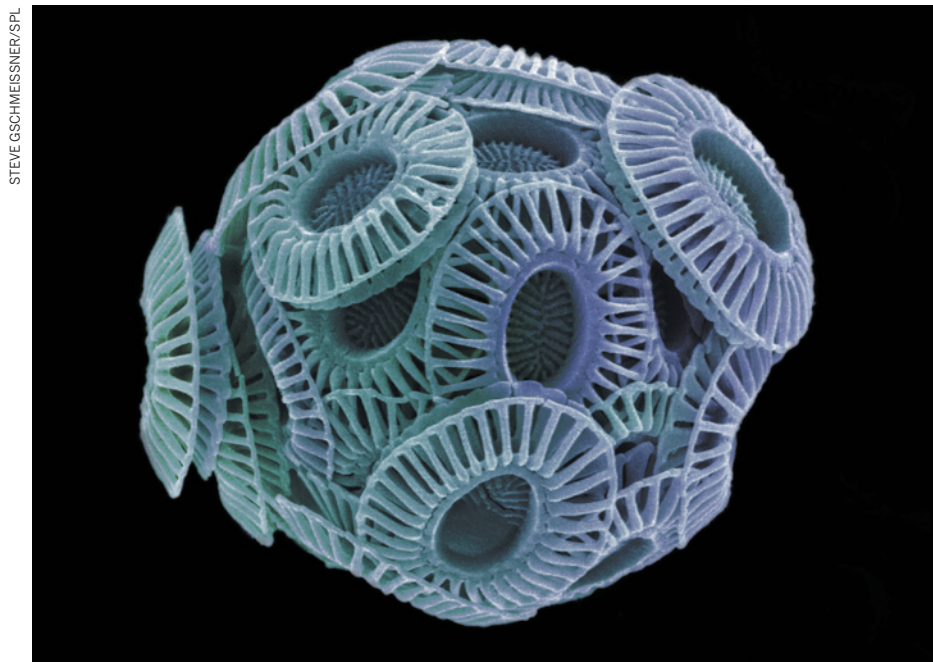


Figure 1 | A proxy for ancient carbon dioxide levels. Coccolithophores are marine algae characterized by an exoskeleton of overlapping plates (coccoliths) composed of calcium carbonate. Bolton and Stoll¹ report that differences in the carbon-isotope composition of large and small coccoliths preserved in the geological record provide information about atmospheric carbon dioxide levels in ancient times.

of the small-celled alga *Emiliania huxleyi*, but Bolton and Stoll's results highlight the need to examine species of a range of sizes. The greater CO₂ sensitivity of large coccolithophores reported by the authors might affect competition among different-sized species, causing small species to outcompete large species at low CO₂ concentrations, and large coccolithophores to proliferate as CO₂ levels rise. The size-dependent concentration of inorganic carbon in coccolithophores could also influence the vertical flux of particulate organic and inorganic carbon in the sea, because cells and coccoliths from large species sink faster than those from smaller species.

Central to Bolton and Stoll's results is the conclusion that, as the concentration of CO₂ in sea water declines, a larger proportion of calcification is supported by CO₂ that enters the cell by diffusion. This CO₂ is converted to bicarbonate inside the cell to compensate for the diversion of pumped bicarbonate to photosynthesis. Coccolithophores might therefore produce less acid during calcification as the concentration of CO₂ rises, because a larger proportion of calcification is fed by bicarbonate (which produces 1 mole of acid per mole of calcium carbonate precipitated) than by CO₂ (which produces 2 moles of acid per mole of calcium carbonate precipitated). As a result, and because coccolithophores account for a large fraction of total calcification in the ocean, the currently expected decrease in the surface ocean's pH may be partially offset as CO₂ levels rise.

Perhaps of greater consequence to the global carbon cycle is how rising CO₂ concentrations will affect calcium carbonate precipitation by

coccolithophores overall. At low concentrations of 200–400 p.p.m., which occurred during past glacial periods and pertain today, calcification seems to decrease as atmospheric CO₂ levels rise⁹. Over a higher concentration range (400–750 p.p.m.), such as that expected during the next 100 years, the evidence relating to calcification trends is inconclusive^{8–10}.

NEUROSCIENCE

Dopamine ramps up

We thought we had figured out dopamine, a neuromodulator involved in everything from learning to addiction. But the finding that dopamine levels ramp up as rats navigate to a reward may overthrow current theories. [SEE LETTER P.575](#)

Yael Niv

Scientific findings typically come in two flavours: explanations for things we already knew occurred but had no idea why, or new phenomena that are clearly important but still mysterious. Howe and colleagues' finding¹, on page 575 of this issue, is of the latter kind — even if we don't yet know what it means, it stands to alter the way we think about dopamine.

Dopamine is a molecule that is broadcast throughout the brain and is involved in processes ranging from decision-making to schizophrenia, as well as most forms of addiction. The authors measured levels of dopamine in the

If, as Bolton and Stoll's findings suggest, the proportion of pumped bicarbonate used for calcification increases at higher concentrations of CO₂, this may partially counteract any suppression of calcification associated with future ocean acidification.

As the concentration of CO₂ in the atmosphere increases over the next 50 to 100 years, the 3-million-year transition from high to low CO₂ levels that Bolton and Stoll conclude occurred some 7 Myr ago will play back in reverse, but 30,000 to 60,000 times faster. Although the effects of this rapid carbonation of Earth's oceans on marine ecology and on the ocean's ability to absorb atmospheric CO₂ are uncertain, Bolton and Stoll provide insight into how a crucial component of marine phytoplankton communities worldwide may respond. ■

John Reinfelder is in the Department of Environmental Sciences, Rutgers University, New Brunswick, New Jersey 08901, USA. e-mail: reinfelder@envsci.rutgers.edu

1. Bolton, C. T. & Stoll, H. M. *Nature* **500**, 558–562 (2013).
2. Beerling, D. J. & Royer, D. L. *Nature Geosci.* **4**, 418–420 (2011).
3. Pagani, M., Freeman, K. H. & Arthur, M. A. *Science* **285**, 876–879 (1999).
4. Pagani, M., Liu, Z., LaRiviere, J. & Ravelo, A. C. *Nature Geosci.* **3**, 27–30 (2010).
5. LaRiviere, J. P. et al. *Nature* **486**, 97–100 (2012).
6. Bidigare, R. R. et al. *Glob. Biogeochem. Cycles* **11**, 279–292 (1997).
7. <http://keelingcurve.ucsd.edu>
8. Bach, L. T. et al. *New Phytol.* **199**, 121–134 (2013).
9. Beaufort, L. et al. *Nature* **476**, 80–83 (2011).
10. Iglesias-Rodriguez, M. D. et al. *Science* **320**, 336–340 (2008).

striatum of rats while the animals ran through mazes for food rewards. The striatum (Fig. 1a) is the area that contains the highest dopamine concentration in the brain. It is involved in action selection at all levels, from choosing which limb to move to selecting a goal to work towards. In a series of elegant experiments, Howe et al. established that dopamine concentration gradually ramps up as rats run towards a reward, and that the slope of the ramps relates to the amount of anticipated reward and the effort required to obtain it.

Why are these dopamine ramps so relevant? Dopamine-secreting (dopaminergic) neurons are special because they are thought to fire in unison, broadcasting a single all-important

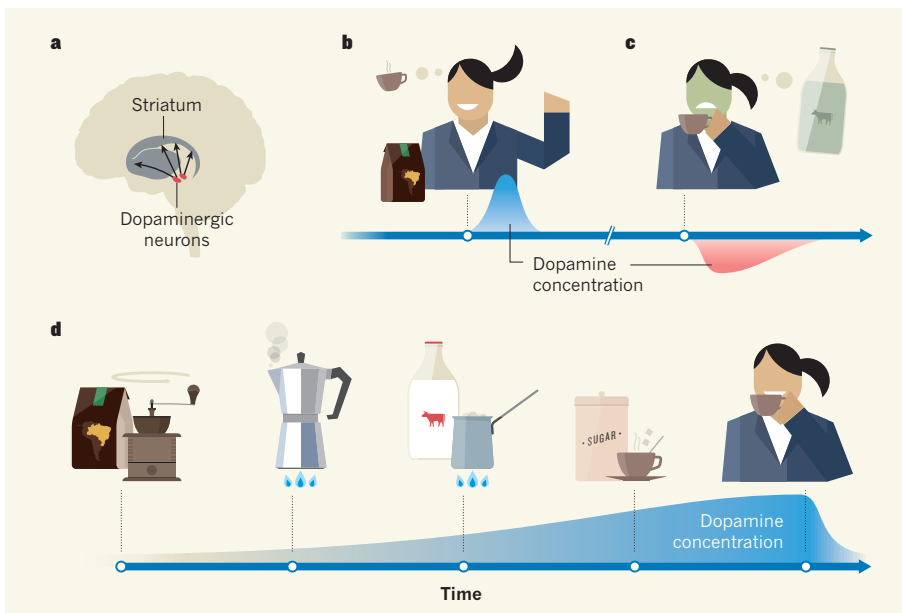


Figure 1 | Rewards and the brain. **a**, Dopaminergic neurons in the midbrain project to all brain areas, most prominently to the striatum (black arrows). These cells fire at a constant rate of 3–5 spikes per second, with occasional phasic bursts or pauses on the occurrence of positive reward prediction errors (discovering that the local supermarket now supplies your favourite coffee beans, **b**) or negative reward prediction errors (sipping your coffee and finding that the milk has gone sour, **c**), respectively. The background (tonic) level of dopamine fluctuates slowly, possibly tracking the average rate of rewards (not shown). **d**, By measuring dopamine concentrations in the striatum of rats navigating mazes, Howe *et al.*¹ reveal a third mode of dopaminergic signalling: when a prolonged series of actions must be completed to obtain a reward (for instance, all the steps it takes to make a cup of coffee), dopamine concentration ramps up gradually, at each point in time signalling the predicted distance from the goal.

message widely. As such, and because dopamine is implicated in a bewildering number of disorders, neuroscientists have long been keen to understand the dopamine signal in the intact brain and how it can be restored when things go wrong. These efforts have led to a well-established theory, which may unfortunately be at odds with these dopamine ramps.

The central theory of dopaminergic firing comes from theoretical neuroscientists who noticed that the firing patterns of dopaminergic neurons bear an eerie resemblance to a key component in computational algorithms of reinforcement learning² called a reward prediction error³. Reward prediction errors quantify ‘surprise’ — the difference between the rewards we expect and those we get in reality⁴.

Imagine you drink coffee routinely. One day, while shopping, you find coffee beans you particularly like, which were not previously available in your town. According to the theory, and verified experimentally in laboratory settings^{5–7}, your dopaminergic neurons will fire to signal a positive reward prediction error due to the increase in your future expected (coffee) rewards (Fig. 1b). Such bursts of dopamine release affect learning in the striatum⁸, strengthening actions that preceded a positive prediction error — you will now be more likely to return to this shop. The flip side, a negative prediction error, occurs when a reward is below expectation — for instance, if you sip your coffee and find

that the milk has gone sour (Fig. 1c).

To the brain, new information that causes you to change your predictions and an actual reward that is at odds with your predictions are equivalent. In both cases, bursts or pauses in dopaminergic firing will notify the brain of the prediction error. Consequently, through dopamine-dependent learning, future predictions will become more accurate, and actions that led to better-than-expected outcomes will become more common.

The prediction-error theory is compelling because it is normative — it explains the role of dopamine within a prescriptive framework of how one should adapt behaviour to earn more rewards. However, it also has shortcomings. For instance, it fails to address dopamine’s effects on action vigour: Parkinson’s disease, caused by the death of dopaminergic neurons, results in slowing down of all actions, and the nickname for amphetamine (which mimics elevated dopamine levels) is ‘speed’.

Luckily, a straightforward extension of the theory fills this gap, suggesting that the background concentration of dopamine (termed tonic dopamine, to differentiate it from the phasic bursts and pauses that signal prediction errors) indicates the overall rate of rewards. Thus, tonic levels of dopamine quantify the cost of time, and should affect how much time we spend on each action⁹. According to this theory, phasic and tonic dopaminergic signalling

convey distinct but related information, affecting learning and vigour, respectively.

So where do the ramps Howe *et al.* observe fit in? They don’t — at least, not in any straightforward way. The ramps do not correspond to anything surprising and so cannot be signalling prediction errors. Because the ramps are predictable, the computational theory of reinforcement learning² dictates that they should disappear through learning. Instead, they persist even when rats have learned to predict the reward. If anything, they then become more informative, with the slope of the ramp differentiating whether the rat expects a small or large reward. Moreover, the ramps do not fit in with ideas about tonic levels of dopamine, because the ramps are too short-lived to be considered tonic and because the rate of reward does not increase as the rat traverses the maze.

What, then, do the ramps tell us about dopaminergic signalling in the brain? Mostly, that we have to go back to the drawing board. Obviously, before radically rethinking the current theory, it should be determined how general these findings are. Others have measured dopamine in tasks requiring a series of actions to obtain rewards, but have not seen reliable evidence of ramps¹⁰. Is this because ramping signals are confined to spatial navigation? Are there no dopamine ramps in our brains when we go about making coffee (Fig. 1d), unless this involves a walk to the nearest café?

Another question pertains to the origin of the ramps. Howe *et al.* used advanced, albeit indirect, methods to measure dopamine concentration in the striatum. By contrast, many other studies track the firing of dopaminergic neurons by recording electrical activity in the midbrain, where the cell bodies lie (Fig. 1a). Such recordings from rats running through mazes have yet to be reported. If similar ramps in neural firing rate were found, this would strengthen the current results. Yet, not finding them would also be revealing, because the relationship between firing of dopaminergic neurons and dopamine release in target areas is complex¹¹. For instance, the ramping activity of nearby striatal neurons (which has been recorded in some^{12,13}, but not all, areas of the striatum) may instigate dopamine release even without firing of the dopaminergic cells¹⁴. Whether this mode of dopamine release exists is not known. Regardless of this, once the boundaries of Howe and colleagues’ finding are established, it will once again be up to the theorists to explain what the ramps actually mean. ■

Yael Niv is at the Princeton Neuroscience Institute and the Psychology Department, Princeton University, Princeton, New Jersey 08544, USA.

e-mail: yael@princeton.edu

1. Howe, M. W., Tierney, P. L., Sandberg, S. G., Phillips,

- P. E. M. & Graybiel, A. M. *Nature* **500**, 575–579 (2013).
2. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* (MIT Press, 1998).
 3. Montague, P. R., Dayan, P. & Sejnowski, T. J. *J. Neurosci.* **16**, 1936–1947 (1996).
 4. Niv, Y. & Schoenbaum, G. *Trends Cogn. Sci.* **12**, 265–272 (2008).
 5. Tobler, P. N., Fiorillo, C. D. & Schultz, W. *Science* **307**, 1642–1645 (2005).
 6. Tobler, P. N., Dickinson, A. & Schultz, W. *J. Neurosci.* **23**, 10402–10410 (2003).
 7. Morris, G., Nevet, A., Arkadir, D., Vaadia, E. & Bergman, H. *Nature Neurosci.* **9**, 1057–1063 (2006).
 8. Reynolds, J. N. J., Hyland, B. I. & Wickens, J. R. *Nature* **413**, 67–70 (2001).
 9. Niv, Y., Daw, N. D., Joel, D. & Dayan, P. *Psychopharmacology* **191**, 507–520 (2007).
 10. Wanat, M. J., Kuhn, C. M. & Phillips, P. E. M. *J. Neurosci.* **30**, 12020–12027 (2010).
 11. Montague, P. R. et al. *J. Neurosci.* **24**, 1754–1759 (2004).
 12. van der Meer, M. A. A. & Redish, A. D. *J. Neurosci.* **31**, 2843–2854 (2011).
 13. van der Meer, M. A. A., Johnson, A., Schmitzer-Torbert, N. C. & Redish, A. D. *Neuron* **67**, 25–32 (2010).
 14. Threlfell, S. et al. *Neuron* **75**, 58–64 (2012).

ASTROPHYSICS

Radioactive glow as a smoking gun

The observation of infrared emission following a short γ -ray burst lends support to the hypothesis that mergers of compact binary systems cause such bursts and produce the heaviest nuclei in the cosmos. [SEE LETTER P.547](#)

STEPHAN ROSSWOG

On 3 June 2013, NASA's Swift satellite detected an intense flash of γ -rays known as a short γ -ray burst. Follow-up observations by the Hubble Space Telescope revealed infrared emission that was present about 9 days after the burst, but had faded away after approximately 30 days. In this issue, Tanvir and colleagues¹ (page 547) propose that this emission was powered by the radioactive decay of heavy elements that were freshly synthesized in the merger of a compact binary system, consisting of either two neutron stars or a neutron star and a black hole. If this interpretation is correct, the observation could have profound consequences for high-energy astrophysics, cosmic nucleosynthesis and the detection of gravitational waves.

There are two flavours of γ -ray burst (GRB), each of different duration. Long bursts (lasting more than about 2 seconds) are produced by the death of a rare breed of massive star, whereas short bursts (less than 2 s) are thought

to result from compact-binary mergers. So far, we know of ten systems containing two neutron stars — extremely densely packed objects that are more massive than the Sun, but only about 12 kilometres in radius, and that consist predominantly of neutrons. As the stars orbit around each other, they emit gravitational waves and slowly spiral inwards until they merge (Fig. 1). Such orbital decays have been observed² and agree remarkably well with the predictions from Einstein's theory of general relativity.

The last moments of the spiral and the final merger are the prime targets for ground-based gravitational-wave laboratories such as LIGO³, VIRGO⁴ or KAGRA⁵. The gravitational-wave signal not only contains information about the physical parameters of the merging system (for example, the masses and spins of the neutron stars), but also carries the imprint of relativistic gravity and of the properties of matter at densities that exceed those inside an atomic nucleus. However, such a signal leaves us essentially blind as to the astronomical environment (for

example, the type of host galaxy) in which the merger occurs. Therefore, accompanying electromagnetic emission that can provide such extra information is highly welcome.

Compact-binary mergers are also interesting from a nucleosynthesis point of view: they eject roughly 1% of a solar mass of neutron-rich matter^{6–8}, and calculations show that this material is entirely made of heavy 'rapid neutron capture' elements^{9–11}. Such elements form through the rapid (as compared with radioactive decays) capture of neutrons on existing atomic nuclei, a mechanism known as the r-process. Generally, adding new neutrons to existing nuclei will not produce stable configurations, thus leading to a competition between further neutron captures and radioactive decays. If a lot of neutrons are present, as in the case of disrupted neutron-star material, very heavy and neutron-rich nuclei can be produced that are far from the 'valley of stability' — the location at which all the stable nuclei reside in the nuclear chart. In the most extreme cases, the neutron capture stops only near the 'neutron dripline', the point at which neutrons are no longer bound and can just 'drip out' of nuclei. Once all available neutrons have been consumed, the unstable nuclei decay towards the valley of stability with a wide variety of radioactive half-lives.

If this rapid neutron capture occurs inside the ejected material and sufficient radioactivity is left when the ejected matter becomes transparent, this should cause a radioactively powered, electromagnetic transient^{12–14}. This phenomenon is often called a macronova or kilonova in the astronomical literature. The delay time between the merger and the peak of the transient emission and its wavelength range depend sensitively on the opacity of the ejected material. Unfortunately, the opacity of such heavy r-process matter is not well known, but recent calculations¹⁵ indicate that it is orders of magnitude larger than that of iron-group-like nuclei. Therefore, macronovae should peak a few days after a merger at infrared wavelengths. The transient reported by Tanvir *et al.* may be the first event of this kind ever to be observed.

If this interpretation is correct, the observation represents a major advance linking



Figure 1 | Supercomputer simulation of the merger of two neutron stars¹⁹. The two stars merge within a few milliseconds and eject around 1% of a solar mass of matter into space. The radioactive decay of this material is likely to have been observed as an infrared transient by Tanvir and colleagues¹. The colour scale denotes the logarithm of temperature in megaelectronvolts. A temperature unit of 1 MeV corresponds to 1.16×10^{10} kelvins.

different fields of research. It would mean that short GRBs are indeed caused by compact-binary mergers — a connection that has been suspected for a long time but never actually proved. It would further imply that an r-process has been caught in the act, or at least its still-glowing ashes. The broad agreement with model calculations suggests that the ejecta do carry roughly 1% of a solar mass. This, in turn, indicates that compact-binary mergers are one of the major cosmic cauldrons in which the heaviest elements are forged. Because macronovae emit their radiation almost the same way in all directions, there is a reasonable chance of observing them in connection with gravitational waves. However, for short GRBs — which are bright but highly beamed and mostly point away from us — the odds of observing them together with gravitational waves are slim. A macronova observation following a claim of gravitational-wave detection would substantially enhance the confidence that the wave was real and not a false alarm, and it could provide an accurate position for its source.

It is worth pointing out, however, that the macronova case is still far from being closed. The data are consistent with the interpretation of a radioactively powered transient, similar to recent model predictions^{16–18}, but we have no spectroscopic evidence that r-process elements have truly been produced. If compact-binary mergers are the ‘central engines’ of short GRBs, they involve a wealth of different physical processes that are not well enough understood to rule out other production channels for infrared emission. To better understand the macronova phenomenon, we need a larger observational sample and, on the theory side, better r-process opacity estimates (current values are based on only a few representative ions) and more detailed calculations of radiation hydrodynamics. Thus, the results must be viewed as tentative for now, but they could be the first step into the multi-messenger era of compact-binary mergers. ■

Stephan Rosswog is in *The Oskar Klein Centre for Cosmoparticle Physics, Department of Astronomy, Stockholm University, AlbaNova, SE-106091 Stockholm, Sweden.* e-mail: stephan.rosswog@astro.su.se

1. Tanvir, N. R. *et al.* *Nature* **500**, 547–549 (2013).
2. Taylor, J. H. & Weisberg, J. M. *Astrophys. J.* **253**, 908–920 (1982).
3. Harry, G. M. *et al.* *Class. Quantum Gravity* **27**, 084006 (2010).
4. Acernese, F. *et al.* *Class. Quantum Gravity* **23**, S63–S69 (2006).
5. Sormiya, K. *Class. Quantum Gravity* **29**, 124007 (2012).
6. Rosswog, S. *et al.* *Astron. Astrophys.* **341**, 499–526 (1999).
7. Bauswein, A., Goriely, S. & Janka, H.-T. *Astrophys. J.* **773**, 78 (2013).
8. Hotokezaka, K. *et al.* *Phys. Rev. D* **87**, 024001 (2013).
9. Freiburghaus, C., Rosswog, S. & Thielemann, F.-K. *Astrophys. J.* **525**, L121–L124 (1999).
10. Goriely, S., Bauswein, A. & Janka, H.-T. *Astrophys. J.*

Lett. **738**, L32 (2011).

11. Korobkin, O., Rosswog, S., Arcones, A. & Winteler, C. *Mon. Not. R. Astron. Soc.* **426**, 1940–1949 (2012).
12. Li, L.-X. & Paczyński, B. *Astrophys. J.* **507**, L59–L62 (1998).
13. Rosswog, S. *Astrophys. J.* **634**, 1202–1231 (2005).
14. Metzger, B. D. *et al.* *Mon. Not. R. Astron. Soc.* **406**, 2650–2662 (2010).
15. Kasen, D., Badnell, N. R. & Barnes, J. Preprint at

<http://arxiv.org/abs/1303.5788> (2013).

16. Barnes, J. & Kasen, D. Preprint at <http://arxiv.org/abs/1303.5787> (2013).
17. Tanaka, M. & Hotokezaka, K. Preprint at <http://arxiv.org/abs/1306.3742> (2013).
18. Grossman, D., Korobkin, O., Rosswog, S. & Piran, T. Preprint at <http://arxiv.org/abs/1307.2943> (2013).
19. Rosswog, S., Piran, T. & Nakar, E. *Mon. Not. R. Astron. Soc.* **430**, 2585–2604 (2013).

GENE REGULATION

Long RNAs wire up cancer growth

The discovery of long non-coding RNAs that control the liaisons between a transcription factor with a key role in prostate cancer and its target genes sheds light on how RNAs dictate information flow in the cell nucleus. SEE LETTER P.598

ADAM M. SCHMITT & HOWARD Y. CHANG

The human genome generates more than 10,000 long non-coding RNA (lncRNA) molecules, yet the functions of only several dozen of these transcripts are known. On page 598 of this issue, Yang *et al.*¹ describe two lncRNAs that bind to, and govern the function of, the androgen receptor — a transcription factor that activates the expression of thousands of genes in response to the hormone androgen*. The authors find that inhibition of these two lncRNAs can block the growth of prostate-cancer cells that are

*This article and the paper under discussion¹ were published online on 14 August 2013.

resistant to hormone therapy owing to a mutation in the androgen receptor.

Androgen signalling has been recognized as the principle driver of prostate-cancer growth since the 1940s, when castration was first shown to slow the progression of prostate cancer². Consequently, androgen-deprivation therapy, which can involve castration, has been the mainstay of treatment for advanced prostate cancer and achieves excellent response rates. However, in most men the cancer becomes resistant within two years of starting this therapy. It is now clear that such resistance arises by various mechanisms that reconstitute androgen-receptor signalling².

A large proportion of lncRNAs associate

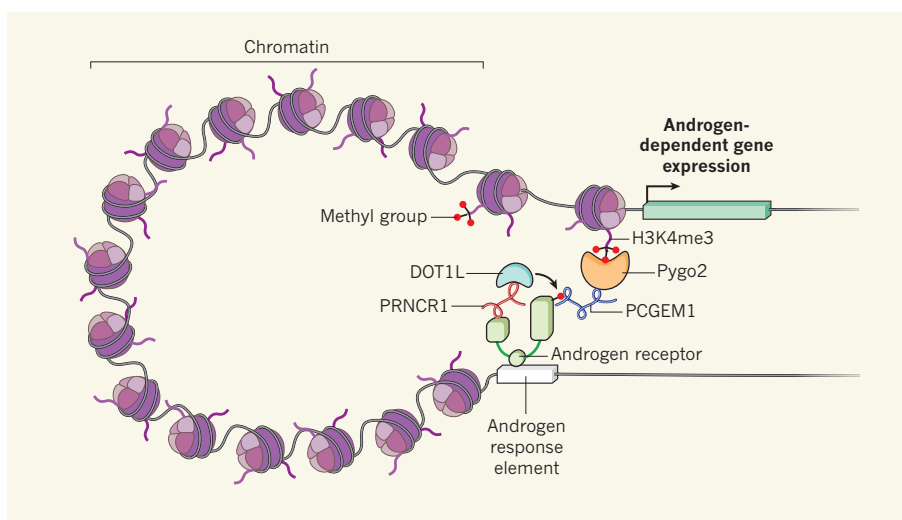


Figure 1 | Long non-coding RNAs and prostate-cancer growth. Yang *et al.*¹ report that two long non-coding RNAs (lncRNAs) — PRNCR1 and PCGEM1 — activate the androgen receptor. Interaction of PRNCR1 with this receptor at androgen-response genomic elements allows recruitment of DOT1L, an enzyme that methylates and so activates the receptor. PCGEM1 can now bind to the active androgen receptor and recruit the enzyme Pygo2, which allows communication between this receptor and its target genes by binding to H3K4me3 chromatin marks in the genes' promoter sequences. Many androgen-receptor target genes have been implicated in prostate-cancer growth¹⁰. (Only one molecule of the androgen-receptor dimer is shown.)

with regulators of chromatin (DNA–protein complexes)³, and several are known to ‘tag’ specific chromosomal regions with particular chromatin marks that modulate the availability of the associated genes for expression⁴. Several prostate-cancer-specific lncRNAs have been identified, and some are associated with distinct subtypes of the disease⁵. In 2012, the US Food and Drug Administration approved the use of the lncRNA PCA3 for the detection of prostate cancer⁶. But, despite the discovery of multiple cancer-related lncRNAs, functions for most of these remain unknown.

Yang *et al.* report that PRNCR1 and PCGEM1, two lncRNAs seen at high levels in many aggressive prostate cancers, enhance androgen-receptor-associated transcriptional programs to promote this cancer’s growth. The authors’ ChIRP analysis (a recently developed method of RNA localization⁷) reveals that these lncRNAs localize to distal androgen-response elements on chromatin, co-localizing with the androgen receptor (Fig. 1). In a surprisingly intricate series of events, PRNCR1 interacts with the acetylated carboxy terminus of the androgen receptor and recruits the enzyme DOT1L to methylate the amino terminus of this receptor; this step is necessary for the subsequent association of the androgen receptor with PCGEM1. PCGEM1 then interacts with the protein Pygo2, which can bind to H3K4me3 — a methyl mark on the chromatin-associated histone H3 protein that is prevalent at gene-promoter sequences.

Yang and colleagues further show that PCGEM1 and Pygo2 facilitate ‘looping’ between enhancer and promoter sequences — a recently recognized role for lncRNAs⁸ — and thus activation of androgen-receptor target genes. Moreover, even a truncated androgen receptor that is always active regardless of the presence of androgen depends on these lncRNAs. And lncRNA inhibition blocks the growth of both androgen-dependent and androgen-independent cancer cells. If these results can be replicated in subsequent studies, they offer a notable advance in the fight against advanced prostate cancer, by highlighting lncRNAs as a class of drug targets.

In addition to their relevance to disease, the current results illuminate several fundamental molecular mechanisms. PRNCR1 and PCGEM1 underscore a new role of RNA interaction with sequence-specific DNA-binding proteins — modification of transcription-factor activity. The liaisons between lncRNAs and transcription factors can program step-wise chemical modifications on transcription factors, gating the successive flow of information from enhancer engagement to target-gene activation.

The insights that these findings provide into how lncRNAs can mediate enhancer–promoter looping are also intriguing. The RNA-mediated recruitment of a protein with intrinsic avidity for a promoter-associated histone mark to

distantly located enhancer elements could stabilize DNA looping and promote communication over three-dimensional space. This would mean that, rather than being simple scaffolds, lncRNAs are more akin to a complex computer circuit board, linking together various disparate molecular components and dictating the logical operation of the system.

Yang and co-workers’ study also raises a plethora of questions. For example, the researchers report a series of apparently specific RNA–protein interactions, the molecular details of which should be investigated. It is not clear whether all of the components involved in these interactions have been identified, and whether similar mechanisms operate for other steroid-hormone receptors or transcription factors.

Moreover, how PRNCR1 and PCGEM1 levels are controlled remains unclear. The authors’ data suggest that defined ranges of PRNCR1 and PCGEM1 levels are crucial for proper androgen signalling, but mechanisms by which lncRNA biogenesis and increases in the levels of these RNAs occur in prostate cancer are not known. With the recent recognition that many lncRNAs are dysregulated in human cancers⁹, understanding how these transcripts are generated and regulated is likely

to take a central place in this field of research. The intricacies of the RNA circuits motivate the development of precise RNA-engineering technologies to reprogram nuclear signalling and for cancer diagnostic and prevention purposes, rather than simply modulating lncRNA levels. ■

Adam M. Schmitt and Howard Y. Chang are at the Howard Hughes Medical Institute and the Program in Epithelial Biology, Stanford University School of Medicine, Stanford, California 94305, USA. A.M.S. is also in the Department of Radiation Oncology, Stanford University School of Medicine.
e-mail: howchang@stanford.edu

1. Yang, L. *et al.* *Nature* **500**, 598–602 (2013).
2. Ryan, C. J. & Tindall, D. J. *J. Clin. Oncol.* **29**, 3651–3658 (2011).
3. Guttman, M. *et al.* *Nature* **477**, 295–300 (2011).
4. Batista, P. J. & Chang, H. Y. *Cell* **152**, 1298–1307 (2013).
5. Prensner, J. R. *et al.* *Nature Biotechnol.* **29**, 742–749 (2011).
6. www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/DeviceApprovalsandClearances/Recently-ApprovedDevices/ucm294907.htm
7. Chu, C., Qu, K., Zhong, F. L., Artandi, S. E. & Chang, H. Y. *Mol. Cell* **44**, 667–678 (2011).
8. Lai, F. *et al.* *Nature* **494**, 497–501 (2013).
9. Brunner, A. L. *et al.* *Genome Biol.* **13**, R75 (2012).
10. Sharma, N. L. *et al.* *Cancer Cell* **23**, 35–47 (2013).

ASTROPHYSICS

A dark cloud unveils its secrets

Massive stars, with masses ten or more times that of the Sun, dominate our Universe. Theories of how these stars form are now being challenged by observations of a collapsing cloud of gas and dust.

JONATHAN C. TAN

Ten thousand light years away, a monster is stirring and a storm is brewing. The ‘monster’ is a swirling cloud of gas and dust with a mass thousands of times greater than the Sun’s. The ‘storm’ is the birth of massive stars from this cloud — within a million years a great outpouring of light and winds from these stars will be sweeping through space, followed eventually by the cataclysm of supernovae. Understanding this phenomenon is important not only for its effects on our Galaxy today, but also because a cluster of Sun-like stars is expected to form alongside these massive stars and we think our own Solar System was born in such an environment¹. Writing in *Astronomy & Astrophysics*, Peretto *et al.*² present new observations of this dark cloud, SDC335 (Fig. 1), including early data from the Atacama Large Millimeter Array

(ALMA) located high in the Chilean Andes, and attempt to answer a basic question: how do massive stars form?

Stars exist with a wide variety of masses, from about one-tenth that of the Sun to at least 100 times more massive. Although we have a well-developed understanding of the formation of low-mass, Sun-like stars³, how massive ones are assembled has been shrouded in mystery. Only one out of every few hundred stars is massive. This rarity means that, in terms of stars undergoing formation, the closest examples of massive stars are much more distant than those of low-mass stars — typically thousands of light years away rather than hundreds. Furthermore, massive-star birth seems to require high densities of interstellar gas, and dust grains that are mixed throughout this gas — galactic smog — obscure our view in the optical and even infrared wavebands. To peer into the

birth clouds of massive stars, astronomers have had to develop telescopes that see fine detail at longer (far-infrared and radio) wavelengths. Owing to the effect of Earth's atmosphere, sometimes these telescopes need to be sent into space, for example the Spitzer and Herschel telescopes, or located at high altitude, like ALMA.

There are several theoretical ideas about how massive stars form. One, known as core accretion^{4,5}, involves an interstellar gas cloud, compressed under its own self-gravity, fragmenting into many 'cores' that have a range of masses. Low-mass stars grow from low-mass cores, massive stars from massive cores. Once a core starts collapsing to form a star, there is relatively limited further accumulation of gas from the surrounding cloud to the core. We think low-mass stars form this way, on the basis of detailed observations of low-mass cores⁶. A second idea, competitive accretion⁷, involves all stars starting to form from low-mass cores but then competing for feeding of additional gas from the surrounding cloud. A star's ultimate mass is determined by its environment: stars grow massive if they are at the centres of clouds undergoing rapid global collapse, which maintains a rich supply of gas to their location. Moreover, massive stars in the process of forming should be seen surrounded by a cluster of lower-mass stars, which eventually comes to dominate the mass of the system and directs the flow of global collapse.

Peretto and colleagues' study bears upon this debate. First, from millimetre-wavelength observations with ALMA of thermal emissions from dust, they derive properties of two massive cores near the cloud centre. These cores are among the most extreme in terms of the concentration of a large mass, of several hundred solar masses, within a small, roughly spherical volume, of radius less than 6,000 astronomical units (AU), where 1 AU is the Earth–Sun distance. If confirmed, these would be very unusual conditions for an interstellar cloud, with 10% of its total mass concentrated in just 0.001% of its volume. However, as the authors describe, mass estimates from millimetre-wavelength dust emission are uncertain, at least by factors of a few. In addition, the small core sizes are not well resolved, but are instead inferred from a slight broadening of their angular profiles compared with the narrowest profile the telescope achieves of a point source. Nevertheless, it seems likely that massive, dense cores have been found that are capable of forming massive stars.

Second, the authors use infrared and far-infrared data from the Spitzer and Herschel telescopes to study the larger-scale cloud. It seems to be threaded by several filaments of gas and dust that extend radially from the centre. Peretto and colleagues argue that this is suggestive of a pattern of global feeding of the cores from the surrounding cloud. To search for motions associated with such feeding, Peretto *et al.* examine spectral-line emission

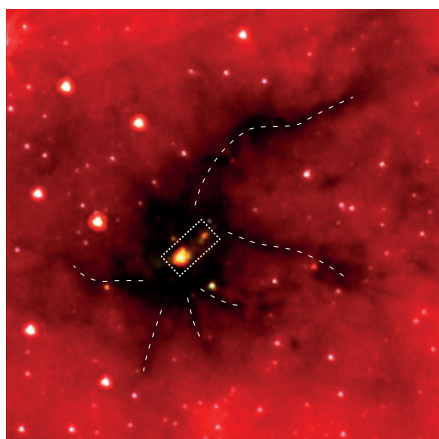


Figure 1 | Collapsing cloud. This infrared image of the SDC335 dark cloud was taken with the Spitzer telescope. Peretto *et al.*² find two massive gas cores (dotted box) near the cloud centre, coinciding with infrared sources, which are likely to be forming massive stars. A web of surrounding filaments (dashed lines) is contracting towards the centre, providing clues to how these cores and stars are forming.

from several molecular tracers, using both ALMA and the Australian Mopra telescope. They detect small Doppler shifts in frequency caused by the motion of these molecules either towards or away from us. From this information, they infer that the cloud is turbulent and also contracting, presumably owing to gravity. The maximum 'free-fall' rate of such collapse occurs if there is no resistance from internal pressure. However, the observed flow along the filaments implies a collapse rate of only about one-tenth of free fall.

Peretto *et al.* conclude that the observation

of global collapse supports the theory of competitive accretion of massive-star formation. Although the evidence for collapse is compelling, this conclusion may be premature. Global collapse alone cannot be used to discriminate between core and competitive accretion, because both involve self-gravitating clouds that must have assembled in this way. Rather, it is the rate of collapse that is likely to be important in determining whether gas that forms massive stars has a chance to first become organized in massive cores or instead is delivered quickly to a central region and there apportioned among a forming cluster, as yet unseen in SDC335. An alternative possible interpretation of the observations is that global collapse is relatively slow and that massive cores have formed and achieved densities much larger than their surroundings, from which they are then effectively decoupled. Observations with a fully operational ALMA will see at least ten times finer detail and will be needed to resolve this debate decisively. ■

Jonathan C. Tan is in the Departments of Astronomy and Physics, University of Florida, Gainesville, Florida 32611, USA.
e-mail: jt@astro.ufl.edu

- Adams, F. C. *Annu. Rev. Astron. Astrophys.* **48**, 47–85 (2010).
- Peretto, N. *et al. Astron. Astrophys.* **555**, A112 (2013).
- Shu, F. H., Adams, F. C. & Lizano, S. *Annu. Rev. Astron. Astrophys.* **25**, 23–81 (1987).
- McLaughlin, D. E. & Pudritz, R. E. *Astrophys. J.* **476**, 750–765 (1997).
- McKee, C. F. & Tan, J. C. *Nature* **416**, 59–61 (2002).
- Bergin, E. A. & Tafalla, M. *Annu. Rev. Astron. Astrophys.* **45**, 339–396 (2007).
- Bonnell, I. A., Bate, M. R., Clarke, C. J. & Pringle, J. E. *Mon. Not. R. Astron. Soc.* **323**, 785–794 (2001).

MICROBIOLOGY

Wealth management in the gut

People whose guts contain a low diversity of bacteria are found to have higher levels of body fat and inflammation than those with high gut–microbial richness. But dietary intervention can help. SEE ARTICLE P.541 & LETTER P.585

SUNGSOON FANG & RONALD M. EVANS

To some extent, each of us is a 'super-organism', composed of our own genome and the collective genetic contribution of the non-pathogenic microorganisms that reside in our guts and other organs. A key feature of this aggregate genome is the dynamic nature of its composition, particularly in response to environmental stressors such as high-fat or low-energy diets. Our bodies can

maintain health in response to acute dietary changes, but chronic nutrient stress can lead to the increased risk of disease, as is seen in the contribution of high-fat diets and sedentary lifestyles to obesity, diabetes, inflammation and heart disease. However, most overweight people have a normal metabolism, and what distinguishes healthy obese individuals from those with metabolic disease is poorly understood. In this issue, Le Chatelier *et al.*¹ (page 541) and Cotillard *et al.*² (page 585)

further our understanding of the role of gut microorganisms in this distinction by identifying a correlation between the genetic 'richness' of the gut microbiome — the collective genome of the resident microorganisms — and health.

A prelude to the current work came from surprising evidence^{3,4} that, in addition to diet and exercise, certain gut bacteria are strongly correlated with, and potentially contribute to, disorders associated with excessive weight. Although all humans host trillions of intestinal bacteria, it was suggested⁵ that the gut microbiome of each person can be assigned to one of three main groups, called enterotypes, which are dominated by species of the genera *Bacteroides*, *Prevotella* or *Ruminococcus*, respectively. Surprisingly, the distribution of enterotypes seems to be independent of sex, age, body-mass index and nationality. The two latest studies built on these ideas to explore microbial gene complexity as a potential marker of metabolic health.

Le Chatelier *et al.* describe differences in the number of gut microbial genes (the bacterial richness) between the microbiomes of non-obese and obese Danish individuals. The authors found that high and low bacterial richness could be distinguished on the basis of the abundance of just a small number of bacterial species. They also found that individuals with low bacterial richness had higher overall levels of body fat (adiposity) and accompanying inflammation-associated characteristics than those with high bacterial richness (Fig. 1). Moreover, among the obese volunteers, those with relatively low microbial richness tended to gain more body weight than those with high microbial richness, suggesting that they harbour inflammation-associated gut microbiota.

These findings suggest that stratification of microbial gene richness is predictive of the metabolic and inflammatory status of the host, and may therefore function as a new biomarker. For example, the data revealed that individuals with a microbiome enriched in *Bacteroides* and some *Ruminococcus* species had an increased incidence of insulin resistance, fatty liver and low-grade inflammation compared with individuals with a microbiota enriched in *Methanobrevibacter* species.

A previous study had found⁶ that a higher dietary intake of animal protein and saturated fat was positively associated with the presence of *Bacteroides* species in the human gut, whereas plant-based nutrition was associated with *Prevotella* species. Cotillard *et al.* show that this is not just an association, but that dietary patterns can predict gut-microbial gene composition. Like Le Chatelier and colleagues, they report that people with low microbial richness (40% of individuals in their study population of obese and overweight French volunteers) had an increased risk of developing metabolic defects compared with those with

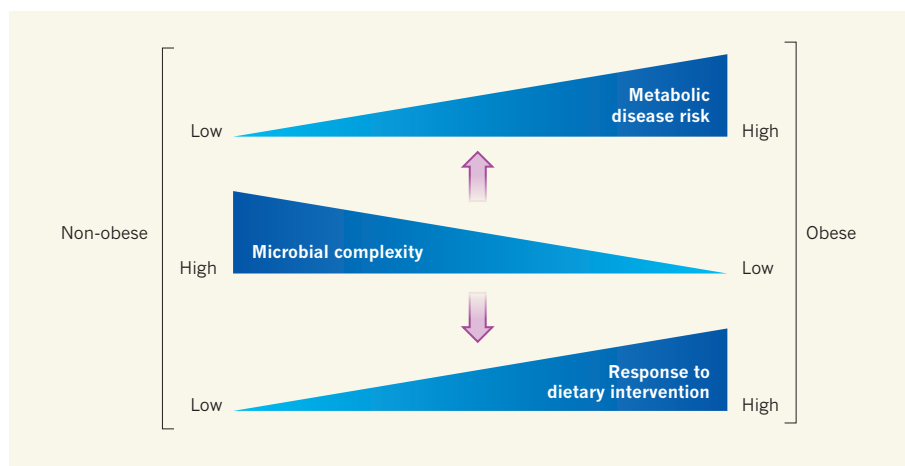


Figure 1 | Benefits of genetic richness. By comparing the complement of microbial genes in the guts of obese and non-obese individuals, Le Chatelier *et al.*¹ show that people with relatively less-complex microbiomes have higher overall body adiposity and more inflammation-associated characteristics, indicating that they are at higher risk of metabolic diseases than people with a greater gut-bacterial richness. Cotillard *et al.*² demonstrate that microbial richness increases, and inflammation decreases, in obese and overweight people with low microbial richness who commence an energy-restricted diet, but that such dietary interventions have little effect in individuals with already high microbial richness.

high microbial richness. The authors then evaluated the effects on these individuals' clinical presentation and microbial composition of 6 weeks of an energy-restricted diet followed by 6 weeks of weight maintenance.

Interestingly, they found that this dietary intervention improved microbial richness and clinical results in individuals who had been classified as having low microbial richness, but that the intervention was significantly less beneficial to those with already high microbial richness (Fig. 1). This suggested that bacterial gene richness is predictive of a person's ability to respond to dietary intervention. Such a concept is promising for clinical research and pharmaceutical drug development, because it opens potential avenues for customizing drugs and reprogramming microbial ecology in individual patients.

Despite these and other advances in our understanding of how bacterial type and diversity influence aspects of health, there remains a debate around the use of three enterotypes as a way of classifying human gut bacteria, and around their physiological significance — some researchers feel that the distinction between microbiomes is less clear cut than simply three dominant types^{6–9}. This controversy matters, because such classifications may eventually affect how a person's relative risk of disease is gauged and what proactive therapeutic options are considered. However, although these groupings need further clarification, it is clear that there are clinically relevant correlations between microbial diversity and physiology. Indeed, the two studies indicate that microbial composition and bacterial gene richness are sufficient to distinguish healthy obese people from obese people with metabolic disease, as well as those who will respond to energy-restriction therapy. This

in turn suggests that categorization of gut microbiota into distinct groups would be an attractive way to use the aggregate microbial genome as a biomarker for predicting the rate of progression in some human disorders, such as inflammatory bowel disease.

These findings also raise the question of whether high microbial gene complexity is merely a reflection of metabolic health or actually offers protection against disease progression. If the former, then microbial conversion from high to low genetic richness would be diagnostic only of a physiological state, as opposed to an indication of risk or a trigger for pathology. Further analysis of the aggregate microbial genomes in multiple populations and testing of microbial transplantation may help to resolve these issues. But in the meantime, be nice to your microbial friends. Although small, they may be powerful allies. ■

Sungsoo Fang and Ronald M. Evans are in the Gene Expression Laboratory, Salk Institute for Biological Studies, La Jolla, California 92037, USA. R.M.E. is also at the Howard Hughes Medical Institute.
e-mails: sfang@salk.edu; evans@salk.edu

1. Le Chatelier, E. *et al.* *Nature* **500**, 541–546 (2013).
2. Cotillard, A. *et al.* *Nature* **500**, 585–588 (2013).
3. Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. *Nature* **444**, 1022–1023 (2006).
4. Turnbaugh, P. J. *et al.* *Nature* **444**, 1027–1031 (2006).
5. Arumugam, M. *et al.* *Nature* **473**, 174–180 (2011).
6. Wu, G. D. *et al.* *Science* **334**, 105–108 (2011).
7. Claesson, M. J. *et al.* *Nature* **488**, 178–184 (2012).
8. Huse, S. M., Ye, Y., Zhou, Y. & Fodor, A. A. *PLoS ONE* **7**, e34242 (2012).
9. Yong, E. <http://dx.doi.org/10.1038/nature.2012.10276> (2012).

Richness of human gut microbiome correlates with metabolic markers

Emmanuelle Le Chatelier^{1*}, Trine Nielsen^{2*}, Junjie Qin^{3*}, Edi Prifti^{1*}, Falk Hildebrand^{4,5}, Gwen Falony^{4,5}, Mathieu Almeida¹, Manimozhiyan Arumugam^{2,3,6}, Jean-Michel Batto¹, Sean Kennedy¹, Pierre Leonard¹, Junhua Li^{3,7}, Kristoffer Burgdorf², Niels Grarup², Torben Jørgensen^{8,9,10}, Ivan Brandslund^{11,12}, Henrik Bjørn Nielsen¹³, Agnieszka S. Juncker¹³, Marcelo Bertalan¹³, Florence Levenez¹, Nicolas Pons¹, Simon Rasmussen¹³, Shinichi Sunagawa⁶, Julien Tap^{1,6}, Sebastian Tims¹⁴, Erwin G. Zoetendal¹⁴, Søren Brunak¹³, Karine Clément^{15,16,17}, Joël Doré^{1,18}, Michiel Kleerebezem¹⁴, Karsten Kristiansen¹⁹, Pierre Renault¹⁸, Thomas Sicheritz-Ponten¹³, Willem M. de Vos^{14,20}, Jean-Daniel Zucker^{15,16,21}, Jeroen Raes^{4,5}, Torben Hansen^{2,22}, MetaHIT consortium†, Peer Bork⁶, Jun Wang^{3,19,23,24,25}, S. Dusko Ehrlich¹ & Oluf Pedersen^{2,26,27,28}

We are facing a global metabolic health crisis provoked by an obesity epidemic. Here we report the human gut microbial composition in a population sample of 123 non-obese and 169 obese Danish individuals. We find two groups of individuals that differ by the number of gut microbial genes and thus gut bacterial richness. They contain known and previously unknown bacterial species at different proportions; individuals with a low bacterial richness (23% of the population) are characterized by more marked overall adiposity, insulin resistance and dyslipidaemia and a more pronounced inflammatory phenotype when compared with high bacterial richness individuals. The obese individuals among the lower bacterial richness group also gain more weight over time. Only a few bacterial species are sufficient to distinguish between individuals with high and low bacterial richness, and even between lean and obese participants. Our classifications based on variation in the gut microbiome identify subsets of individuals in the general white adult population who may be at increased risk of progressing to adiposity-associated co-morbidities.

Modern living with a sedentary everyday life, a constant boom of easily accessible and energy-dense food, and exposure to additional ‘obesogenic’ environmental factors together with extended life expectancy has resulted in an epidemic of metabolic disorders characterized by a core of excessive body fat accumulation. Projection estimations predict that, on a global scale, cases will rise from 400 million obese adults in 2005 to more than 700 million in 2015, and this trend will continue towards 2030 (refs 1, 2). Some individuals seem to be more susceptible to the obesogenic environment of modern living than others, suggesting an important inherited component, supported by several twin, family and adoption studies, with heritability estimates ranging from 40% to 70% (refs 3–5). Studies of variation in the human genome have so far resulted in the discovery of more than 50 validated genome-wide significant loci associated with overall adiposity and body composition^{6–12}. Yet, despite a reasonable number of obesity susceptibility variants identified, the proportion of explained genetic variance of body mass index (BMI) remains low, that is, a few per cent (ref. 6). Emerging evidence suggests, however, that variation in our ‘other genome’—the collective genome of the microorganisms inhabiting our body, known as the

microbiome—may have an even greater role than human genome variation in the pathogenesis of obesity given its direct interaction with environmental factors. Recent studies show that the human gut microbiota may be altered in obese relative to lean individuals, even if inconsistent changes have been reported. An increase in the phylum Firmicutes and a decrease in Bacteroidetes associated with obesity was observed in some^{13,14}, but not all, studies¹⁵, with the inverse also reported¹⁶. An increase of Actinobacteria in obese individuals was also reported¹⁷. Mouse gut microbiota obesity-related alterations are characterized by changes in the Firmicutes to Bacteroidetes ratio, which is increased in the obese animals^{18,19}. These changes are probably not a mere consequence of obesity, because the obese phenotype can be transmitted by gut microbiota transplantation in mice, indicating that gut microbial populations may have an active role in obesity pathogenesis^{20,21}. Establishment of a catalogue of bacterial genes from the human gut²² encouraged us to address the hypothesis that variation in the gut microbiome at gene and species levels defines subsets of individuals in the adult population who are at increased risk of obesity-related metabolic disorders.

¹INRA, Institut National de la Recherche Agronomique, US1367 Metagenopolis, 78350 Jouy en Josas, France. ²The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, DK-2200 Copenhagen, Denmark. ³BGI-Shenzhen, Shenzhen 518083, China. ⁴Department of Structural Biology, VIB, Pleinlaan 2, 1050 Brussels, Belgium. ⁵Department of Bioscience Engineering, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium. ⁶European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany. ⁷School of Bioscience and Biotechnology, South China University of Technology, Guangzhou 510006, China. ⁸Research Centre for Prevention and Health, Glostrup University Hospital, DK-2900 Glostrup, Denmark. ⁹Faculty of Health and Medical Sciences, University of Copenhagen, DK-2200 Copenhagen, Denmark. ¹⁰Institute of Public Health, Faculty of Medicine, University of Aalborg, DK-9100 Aalborg, Denmark. ¹¹Department of Clinical Biochemistry, Vejle Hospital, DK-7100 Vejle, Denmark. ¹²Institute of Regional Health Research, University of Southern Denmark, DK-8200 Odense, Denmark. ¹³Center for Biological Sequence Analysis & Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark. ¹⁴Laboratory of Microbiology, Wageningen University, 6710BA Ede, The Netherlands. ¹⁵Institut National de la Santé et de la Recherche Médicale, U872, Nutrimique, Équipe 7, Centre de Recherches des Cordeliers, 75006 Paris, France. ¹⁶Université Pierre et Marie-Curie-Paris VI, 75006 Paris, France. ¹⁷Assistance Publique-Hôpitaux de Paris, Institute of Cardiometabolism and Nutrition, CRNH-Ile de France, Pitié-Salpêtrière, 75013 Paris, France. ¹⁸INRA, Institut National de la Recherche Agronomique, UMR 14121 MICALIS, 78350 Jouy en Josas, France. ¹⁹Department of Biology, Ole Maaløes Vej 5, University of Copenhagen, DK-2200 Copenhagen, Denmark. ²⁰Department of Bacteriology and Immunology, University of Helsinki, FIN-00014 Finland. ²¹Institut de Recherche pour le Développement, UMI 209, Unité de modélisation mathématique et informatique des Systèmes Complexes, F-93143 Bondy, France. ²²Faculty of Health Sciences, University of Southern Denmark, DK-8200 Odense, Denmark. ²³King Abdulaziz University, Jeddah 21589, Saudi Arabia. ²⁴Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, DK-2200 Copenhagen, Denmark. ²⁵Center for Sequencing Aarhus University, DK-8000 Aarhus C, Denmark. ²⁶Hagedorn Research Institute, DK-2820 Gentofte, Denmark. ²⁷Institute of Biomedical Science, Faculty of Health and Medical Sciences, University of Copenhagen, DK-2200 Copenhagen, Denmark. ²⁸Faculty of Health, Aarhus University, DK-8000 Aarhus, Denmark.

*These authors contributed equally to this work.

†A list of authors and affiliations appears at the end of the paper.

The abundance of known intestinal bacteria can be assessed by the mapping of a large number of sequencing reads from total faecal DNA onto a reference set of their genomes²³. This approach, which we term quantitative metagenomics, was extended here to assess the abundance of genes from the reference catalogue in a cohort of 292 non-obese and obese individuals.

Bimodal distribution of microbial genes

Comparison of gene number across the total study sample of 292 individuals showed a bimodal distribution of bacterial genes (Fig. 1a and Supplementary Table 1). A similar distribution was detected in obese French individuals using a different sequencing technology²⁴. As the number of genes detected had some dependence on the number of matched reads (Supplementary Fig. 1), we downsized the data set to 11 million reads, thus excluding 15 individuals and the bimodal distribution was again observed (Fig. 1b). We term hereafter the individuals with <480,000 genes 'low gene count' (LGC) and others 'high gene count' (HGC). They had, on average, 380,000 and 640,000 genes, a difference of some 40% and harboured less or more rich microbiota, respectively, as shown by scoring several single copy marker genes (Supplementary Fig. 2). Human intestinal tract chip (HITChip) analysis²⁵, based on the widely accepted 16S ribosomal DNA phylogenetic marker, confirmed the bimodal distribution and the difference of richness of microbial communities between the LGC and HGC individuals (Supplementary Fig. 3 and Supplementary Table 1).

Low richness of gut microbiota has been reported in patients with inflammatory bowel disorder (IBD)^{22,26,27}, elderly patients with inflammation²⁸ and in obese individuals¹⁷, but the differences of richness within these groups or among non-obese individuals were not previously detected. As the composition of the gut microbiota seems to be rather stable over long periods of adulthood²⁹, its richness may well be a characteristic feature of an individual. In mice, the richness seems to be affected by repeated antibiotic treatments³⁰, and host genetics could also have a role³¹. Also notable diversity differences were observed between the urban US population and rural populations from two developing countries³². Further studies, focusing specifically on the

richness of the gut microbiota across broad cohorts, might help to determine the causes for its variation.

Microbial species of LGC and HGC groups

The differences in gene numbers indicate that the LGC and HGC individuals contain different microbial communities. To assess the difference in phylogenetic composition between the two, we combined reference genome mapping with gene abundance data at the phylum, genus and species level.

We first examined the general phylogenetic composition at higher taxonomic levels based on sample-wise rarefied read abundances that were mapped on publicly available reference genomes and binned at the genus and phylum level. Forty-six genera differed significantly in abundance between the HGC and LGC individuals (Supplementary Table 2). Although *Bacteroides*, *Parabacteroides*, *Ruminococcus* (specifically *R. torques* and *R. gnavus*), *Campylobacter*, *Dialister*, *Porphyromonas*, *Staphylococcus* and *Anaerostipes* were more dominant in LGC, 36 genera, including *Faecalibacterium*, *Bifidobacterium*, *Lactobacillus*, *Butyrivibrio*, *Alistipes*, *Akkermansia*, *Coprococcus* and *Methanobrevibacter*, were significantly associated with HGC. For 33 of these, genera probes were present on the HITChip and their abundance was determined (Supplementary Table 2); the correlation with read mapping was very high (Pearson's *r* of 0.85 and 0.90 in HGC and LGC individuals, respectively). At the phylum level, this phylogenetic shift resulted in a higher abundance of Proteobacteria and Bacteroidetes in LGC individuals versus increased populations of Verrucomicrobia, Actinobacteria and Euryarchaeota in HGC individuals (Supplementary Fig. 4).

Next, we studied the specific species that were differentially abundant between LGC and HGC individuals. In this approach, we identified the genes that were significantly different between the LGC and HGC individuals by the Wilcoxon rank-sum test, comparing 204 (70% of total) randomly chosen individuals 30 times. We similarly compared 126 'extreme' individuals, containing <400,000 genes or >600,000 genes. A total of 120,723 genes were found in all 60 tests at $P < 0.0001$ and were analysed further.

We searched for genes from the same species by comparison with all sequenced genomes (Supplementary Materials). At a threshold of 95% identity (the species-level cut-off²³) over at least 90% of the gene length, 10,225 genes (8.5%) were assigned to a total of 97 genomes representing some 73 species (Supplementary Table 3). However, a vast majority (93.4%) belonged to only 9 species, which varied significantly in abundance between the LGC and HGC individuals, as illustrated in Fig. 2a (upper part), where the presence and abundance of 50 arbitrarily chosen 'tracer' genes from each species in the individuals of the cohort is displayed. As expected, these genes have a sharply bimodal distribution: 71% of individuals had either all or none of the genes from a species and thus harboured or lacked that species, at the present depth of analysis. The first five species were more frequent in LGC individuals, whereas the last four species were more frequent in the HGC group.

Taken together, our analyses highlight the contrast between the distribution of anti-inflammatory species, such as *Faecalibacterium prausnitzii*^{33,34}, which are more prevalent in HGC individuals and potentially pro-inflammatory, *Bacteroides* and *R. gnavus*³⁵, associated with IBD^{36,37} and found to be more frequent in LGC individuals.

However, a vast majority (>90%) of the 120,723 genes with significantly differing abundances between the LGC and HGC individuals could not be assigned to a known bacterial genome. These genes must also belong to bacterial species that are present at different abundances in the two groups of individuals. We thus attempted to cluster the genes from the same species using a gene abundance-based approach.

We proposed that the genes of a given bacterial species should be present at a similar abundance in an individual but should display large variations across a cohort, as species abundance is known to vary immensely among individuals (10–10,000-fold)²². The genes that vary in abundance in a coordinated way are thus likely to be from the same

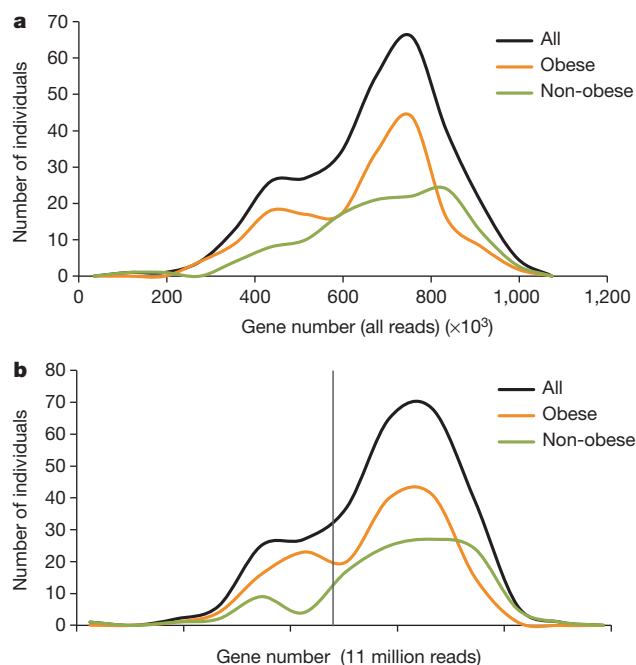


Figure 1 | Distribution of low and high gene count individuals ($n = 292$). **a**, Gene counts from all uniquely matched reads. **b**, Gene counts adjusted to 11 million uniquely mapped reads per individual. Vertical line indicates the threshold of the LGC and the HGS individuals; the observed bimodal distribution was not statistically significant by the dip-test.

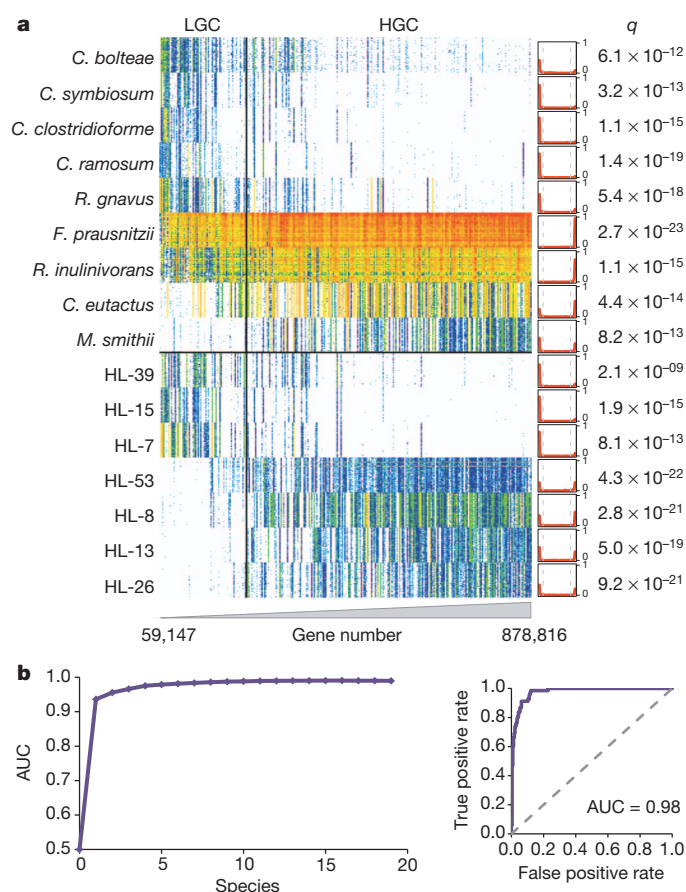


Figure 2 | Bacterial species with different distribution among 292 HGC and LGC individuals. **a**, Presence and abundance of 50 'tracer' genes for representative species differentially abundant in LGC and HGC groups; Mann-Whitney probability (q value, FDR adjusted⁴⁰) is given. Genes are in rows, frequency is indicated by colour gradient (white, not detected; red, most abundant); individuals, ordered by increasing gene number, are in columns, proportion of the tracer genes in individuals is shown on the right. Top, known species; bottom, unknown species. *C. bolteae*, *Clostridium bolteae*; *C. clostridioforme*, *Clostridium clostridioforme*; *C. eutactus*, *Coprococcus eutactus*; *C. ramosum*, *Clostridium ramosum*; *C. symbiosum*, *Clostridium symbiosum*; *F. prausnitzii*, *Faecalibacterium prausnitzii*; *M. smithii*, *Methanobrevibacter smithii*; *R. gnavus*, *Ruminococcus gnavus*; *R. inulinivorans*, *Roseburia inulinivorans*. **b**, Left, AUC values for the best combinations of species in a ROC analysis. Right, AUC for the best combination of four species.

species. We tested this hypothesis for the 10,225 taxonomically assigned genes that differ significantly between LGC and HGC individuals, by computing the Spearman correlation coefficients for each gene with all the other genes and grouping those that were correlated above a given threshold. A large majority (8,125; 79.4%) clustered into only 8 groups that included the 9 most highly represented species shown in Fig. 2 (*Clostridium bolteae* and *C. clostridioforme* genes were in the same group). The specificity and the sensitivity of clustering were very high (average of 97.8% and 91.8%, respectively, for 7 homogeneous groups), indicating that the approach is efficient and can be used to cluster all the significantly different genes.

In total, 76,564 genes (63% of 120,723) were grouped into 1,440 clusters of two or more genes at a threshold of $\rho > 0.85$, used to favour the specificity of clustering. Some 58 clusters contained ≥ 75 genes and included 90% of the genes; 52 contained genes from previously unknown species (Supplementary Table 4). Genes from a cluster originated from the same species in most cases, as shown by (1) coherence of the BLASTP taxonomic assignments; (2) homogeneity of abundance and abundance variation; (3) homogeneity of tetramer composition³⁸; and (4) significant physical linkage (Supplementary

Figs 5–8). This conclusion was further supported by the correlation of the abundance of the 16S rRNA gene sequences represented on the HITChip for 27 of the 58 clusters (Supplementary Table 4; it is possible that the HITChip resolution of closely related genomes may have been insufficient or that the corresponding 16S sequences were lacking for 31 clusters). We conclude that the clustering procedure grouped the genes of the same species, even if in some cases genes from more than a single species were grouped (*C. bolteae* and *C. clostridioforme*, or probably prophages present in genomes from several species; Supplementary Fig. 9). A similar approach was applied in a recent paper, published while the present manuscript was under revision³⁹.

Analyses of the 50 genes from the clusters with known taxonomic assignment have shown that they are present on cognate genomes only and on all cognate genomes (Supplementary Table 5). By extrapolation, we suggest that the same holds true for unknown groups and that the cluster genes can be used as tracers for the species they derive from. Average abundance of the tracer genes was thus equated to the average abundance of the cognate species.

Distribution of unknown species across LGC and HGC individuals of the cohort was clearly biased (Fig. 2a, lower part, and Supplementary Fig. 10). Genes for 10 of the species and the *Bacteroides* genus were present on the metagenomic arrays (Methods); in all cases the HGC/LGC bias found by sequencing was also detected by the arrays (Fig. 2 and Supplementary Table 4). BLASTP analysis indicated that 37% and 92% of the clusters had at least 80% of the genes with the same taxonomy at a genus and phylum level, respectively, a value similar to that observed for all clusters analysed in this paper (Supplementary Fig. 11). HITChip and BLASTP taxonomic assignments were not fully overlapping but whenever both were available they were congruent; when combined, up to 63% of the species could be assigned to a genus. However, there was no obvious stratification of the species prevalent in LGC and HGC individuals at this taxonomic level (Supplementary Table 4).

To test whether LGC and HGC individuals could be distinguished by the bacterial species they contain, we performed a receiver operating characteristic (ROC) analysis. First, we estimated the abundance of 58 species that were significantly different in abundance between LGC and HGC individuals (Supplementary Table 6). For each individual, we used these values to compute a score, named decisive bacterial abundance (DBA) score, equal to the sum of abundances of the species more frequent in HGC individuals subtracted by the sum of the abundances of species more frequent in LGC individuals. The DBA scores were calculated exhaustively for all combinations of up to 19 species and were used in the ROC analysis; the area under curve (AUC) values for the best combinations are shown in Fig. 2b, left. The best combination of four species gave an AUC value of 0.98 (Fig. 2b, right); in a tenfold cross-validation test⁴⁰ with 90% of randomly chosen individuals the AUC value of 0.976 ± 0.02 (mean \pm s.d.) was obtained for the groups of the remaining 10%, indicating the robustness of the analysis. Selection of the most distinctive species on the entire cohort does not seem to lead to a significant over-fit, as the algorithm established with one cohort gives comparably high AUC values with an unrelated cohort²⁴. Future work, searching for correlations of gene abundances and gene counts without separation of individuals into the LGC and HGC groups, may allow identifying additional species that explain variation of gene numbers.

Microbial metabolism of LGC and HGC groups

Through the functional annotation of the reference gene catalogue to KEGG Orthology (KO) groups, abundances of KO groups were determined for the LGC and HGC individuals as in ref. 22. Using the enzyme annotations of the different KO groups, 51 manually defined gut metabolic pathway modules (Supplementary Table 7) differed significantly in abundance between both groups (Methods). LGC individuals had a higher abundance of peroxidase, catalase and TCA modules, suggesting increased capacity to handle exposure to oxygen/oxidative

stress. Furthermore, the genomic potential for production of metabolites with possible deleterious effects on host health (among which pro-carcinogens)—including modules for β -glucuronide degradation, degradation of aromatic amino acids, and dissimilatory nitrate reduction—was significantly higher in LGC participants. Many of the significantly increased modules could be due to the increased abundance of *Bacteroides* spp. (for example, pectin degradation). By contrast, HGC individuals were characterized by a potentially increased production of organic acids—including lactate, propionate and butyrate—combined with a higher hydrogen production potential. Concerning hydrogen removal, a shift from a methanogenic/acetogenic ecosystem in HGC individuals towards a sulphate-reducing one in LGC individuals might take place. The functional capacity of the microbiota in LGC and HGC individuals, when combined with the phylogenetic signal, leads to several interesting observations: in the former group we see (1) a reduction of butyrate-producing bacteria; (2) increased mucus degradation potential combined with a decreased *Akkermansia* to *R. torque/gnavus* ratio; (3) reduced hydrogen and methane production potential combined with increased hydrogen sulphide formation potential; (4) an increase in *Campylobacter/Shigella* abundance; and (5) an increased potential to manage oxidative stress (peroxidase). Overall, this suggests that LGC individuals harbour an inflammation-associated microbiota (Fig. 3).

Phenotypes of the HGC and LGC groups

Characteristics of study materials are given in Supplementary Table 8. We performed an anthropometric and biochemical phenotyping of multiple interrelated features of LGC and HGC individuals, and identified significant differences between them at a false discovery rate (FDR)⁴¹ of up to 10% (Table 1 and Supplementary Table 9). This value was used to avoid missing significant associations; a less stringent level, up to 25%, was chosen in a recent and comparable study design⁴². The LGC individuals, who represented 23% of the total study population, included a significantly higher proportion of obese participants (Fig. 4a; the difference is significant for men and a trend is detected for women), and were as a group characterized by a more marked adiposity, as reflected by an increase in fat mass percentage and body weight (Table 1). The adiposity phenotype of LGC people was associated with increased serum leptin, decreased serum adiponectin, insulin resistance, hyperinsulinaemia, increased levels of triglycerides and free fatty acids, decreased HDL-cholesterol and a more marked inflammatory phenotype (increased highly sensitive C-reactive protein (hsCRP) and higher white blood cell counts) than seen in HGC individuals (Table 1). We further

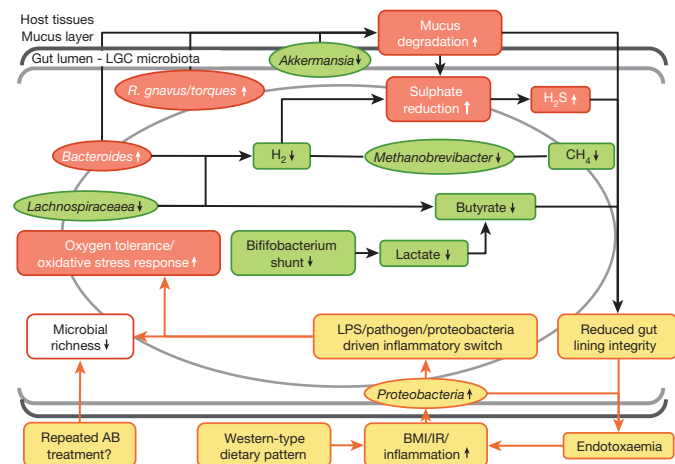


Figure 3 | Functional and phylogenetic shifts in the LGC microbiome. Top, observed increase (red) or decrease (green) of functions and phylogenetic groups. Bottom, potential drivers (yellow) of inflammation related to decreased richness. Left, antibiotic-mediated perturbation of the richness; Right, proteobacterial lipopolysaccharide-mediated perturbation of the richness. AB, antibiotic; IR, insulin resistance.

tested the significance of our observations by treating the gene counts as a continuous variable and examining its correlation with the anthropometric and biochemical variables. All but two (BMI and weight) of the observed differences between LGC and HGC individuals were found significantly associated with the gene counts (Table 1). Together, these analyses suggest that the LGC individuals are featured by metabolic disturbances known to bring them at increased risk of pre-diabetes, type-2-diabetes and ischaemic cardiovascular disorders^{43,44}. Similar abnormalities were found in the accompanying paper²⁴.

We propose that an imbalance of potentially pro- and anti-inflammatory bacterial species triggers low-grade inflammation and insulin resistance (Fig. 3). In parallel, we suggest that an altered gut microbiota of LGC individuals induces the noted increase in levels of serum fasting induced adipose factor (FIAF, also known as ANGPTL4), eliciting an increased release of triglycerides and free fatty acids (Table 1), as evidenced by studies in rodent models^{45–47}. Broad spectra antibiotics may improve glycaemic regulation and change the hormonal, inflammatory and metabolic status in obese mice. Possibly, the reduction of diversity mediated by the antibiotic treatments has a different effect in mice⁴⁸ and man, or, alternatively, is counterbalanced by the restoration of the pro- and anti-inflammatory species balance, which may have been altered in the obese animals. Antibiotic use in early childhood, which may have affected the richness, led to an increased risk of overweight⁴⁹.

Interestingly, obese LGC individuals gained on average significantly more weight than HGC individuals during the past 9 years (Fig. 4b); the BMI change was significant without and with linear adjustment for baseline BMI and age. We searched for species associated to the BMI change among the 58 species that differed significantly between LGC and HGC individuals (Supplementary Table 6) and found eight (Fig. 4c). The average weight gain of individuals with the lowest or undetectable levels of a species was in all cases greater than that of their counterparts with the highest species levels; all eight species were more abundant in HGC than in LGC individuals. These species may therefore protect against weight gain. All but one (*Methanobrevibacter smithii*) lack species-level taxonomic assignment, but four could be assigned at a genus level (*Anareotruncus colihominis*, *Butyrivibrio crossotus* and *Faecalibacterium*; Supplementary Table 4). All are butyrate producers, in agreement with the general tendency of lower butyrate producers among the LGC individuals.

Gut microbes of lean and obese individuals

We also attempted to assess the difference in bacterial species between the lean (BMI < 25 kg m⁻², *n* = 96) and obese (BMI > 30 kg m⁻², *n* = 169) individuals by the approach used for LGC/HGC individuals (Methods). Only 15,894 significantly different genes (*P* < 0.05) were found, indicating that the gut microbiota of lean and obese individuals differs less than that of the LGC and HGC individuals. The genes were attributed to 18 species by the covariance-based clustering (Supplementary Fig. 12 and Supplementary Tables 10 and 11). To test whether lean and obese individuals can be distinguished by these species, we carried out an exhaustive ROC analysis, with tenfold cross validation (Supplementary Fig. 12). The best AUC, of 0.78, was reached with nine species. This accuracy, albeit lower than that for the separation of LGC and HGC individuals, is substantially better than an AUC of 0.58, achieved by ROC analysis of 32 human genome loci associated with adiposity measures⁶. Accordingly, we suggest that the obesity-associated signal in the human gut microbiome may be much stronger than that presently known in the human genome. This view is supported by efficient discrimination of lean and obese individuals in a previous study, in which an AUC of 0.88 was reached with a combination of 50 16S-defined operational taxonomic units (OTUs), separated at the 92% homology level⁵⁰.

Discussion

Contemporary lifestyle is associated with a tide of metabolic abnormalities characterized by a core of excessive body fat accumulation. However, obesity is not just obesity. Some obese individuals seem to

Table 1 | Characteristics of 292 participants stratified by low and high gene counts

	LGC	HGC	LGC versus HGC		Gene count	
			P	q	P	q
N (men/women)	68 (23/45)	224 (113/111)			277 (133/144)*	
Age (years)	55 (50–62)	57 (50–61)	0.86	0.89	0.81	0.84
BMI (kg m ⁻²)	32 (29–34)	30 (23–33)	0.035	0.059	0.11	0.18
Weight (kg)	95 (75–103)	86 (71–102)	0.019	0.037	0.12	0.18
Whole body fat (%)	37 (29–42)	31 (25–39)	0.0069	0.022	0.0024	0.014
S-insulin (pmol l ⁻¹)	50 (35–91)	44 (26–66)	0.0095	0.023	0.0052	0.018
HOMA-IR	1.9 (1.2–3.3)	1.6 (0.9–2.6)	0.012	0.027	0.0059	0.018
P-triglycerides (mmol l ⁻¹)	1.32 (0.97–1.76)	1.15 (0.82–1.57)	0.0014	0.013	0.00073	0.0062
P-free fatty acids (mmol l ⁻¹)	0.55 (0.39–0.70)	0.48 (0.35–0.60)	0.014	0.029	0.00042	0.0062
P-ALT (U l ⁻¹)	20 (14–30)	19 (15–26)	0.22	0.31	0.029	0.06
S-leptin (μg l ⁻¹)	17.0 (6.7–32.6)	8.3 (3.4–26.4)	0.0036	0.019	0.00058	0.0062
S-adiponectin (mg l ⁻¹)	7.5 (5.5–12.9)	9.6 (6.7–13.7)	0.006	0.022	0.016	0.036
B-leucocytes (10 ⁹ l ⁻¹)	6.4 (5.2–7.8)	5.6 (4.8–6.9)	0.0021	0.014	0.0026	0.014
B-lymphocytes (10 ⁹ l ⁻¹)	2.1 (1.6–2.3)	1.8 (1.5–2.1)	0.00082	0.012	0.0037	0.015
P-hsCRP (mg l ⁻¹)	2.3 (1.1–5.7)	1.4 (0.6–2.7)	0.00088	0.012	0.0038	0.015
S-FIAA (μg l ⁻¹)	88 (72–120)	78 (60–101)	0.0047	0.021	0.0088	0.023

Descriptive data are reported as median and interquartile range. To test for differences between the HGC and LGC group, a linear model adjusting for age and sex (*P*) was applied. In the analysis of plasma triglycerides, treatment for lipid lowering medications was added as a covariate to the linear model. Benjamini–Hochberg method was used for multiple testing corrections setting the FDR at 10% (*q*). A similar model was applied to test for associations with gene counts. The P-, S- and B- prefixes denote plasma, serum and blood. ALT, alanin aminotransferase; HOMA-IR, homeostatic model assessment of insulin resistance.

**n* = 277 owing to downsizing of the reads to 11 million.

have a benign prognosis, whereas others progress to co-morbidities such as type-2 diabetes, ischaemic cardio- and cerebrovascular disorders, and non-alcoholic liver disorders. It is also recognized that human obesity in the context of pathogenesis, pathophysiology and therapeutic responsiveness is a heterogeneous condition. The present report provides evidence that studies of alterations in our other genome—the microbial gut metagenome—may define subsets of adult individuals with different metabolic risk profiles and thereby contribute to resolve some of the heterogeneity associated with adiposity-related phenotypes.

We demonstrate that an almost perfect stratification of LGC and HGC individuals can be achieved with very few bacterial species, suggesting that simple molecular diagnostic tests, based on our other genome, can be developed to identify individuals at risk of common

morbidities. Therefore, focus on our other genome may spearhead development of stratified approaches for treatment and prevention of wide-spread chronic disorders. Beyond metabolic dysfunctions, low-grade inflammation is associated with a plethora of chronic diseases. Whether a low gut bacterial richness is common to many or even all of those, as already reported for IBD^{22,26,27}, could be revealed by exploring gut microbiota at a deep metagenomic level in a broad variety of these afflictions.

METHODS SUMMARY

Informed consent was obtained from all 292 volunteers from the Ethical Committees of the Capital Region of Denmark before participation in the study. All individuals were examined after an overnight fast with blood sampling and anthropometric measurements. Faeces samples were collected from all volunteers and frozen immediately; DNA was extracted and sequenced as described²². Gene frequency profiles were established by matching the sequencing reads from an individual sample onto the gut bacterial catalogue²², genes significantly different by frequency between the groups of individuals were identified by Wilcoxon rank-sum test. Genes were clustered by abundance covariance at a selected Spearman correlation coefficient. Taxonomic gene assignments were carried out using BLASTN against an in-house reference catalogue of 3,340 complete and draft microbial genomes downloaded from NCBI databanks at a cut-off of 95% identity²³. Functional annotation was carried out by BLASTP against the eggNOG and KEGG databases as reported²². Abundance of a given species in each individual was estimated as a mean abundance of 50 tracer genes for that species, a value very close to that of the mean abundance of all genes of that species. ROC analysis was based on a combination of bacterial species. For each combination, only a single model was considered, based on the DBA score, computed as the sum of abundances of the species more frequent in one group of individuals subtracted by the sum of the abundances of species more frequent in another group. To select the best models, we used the cross-validated AUC criterion, well adapted to classification models for binary outcome data. Association of microbial composition and metabolic traits were analysed as cross-sectional or continuous variable, upon appropriate normalization; the Benjamini–Hochberg⁴¹ method to correct for multiple testing was used, setting the FDR at 10%.

Full Methods and any associated references are available in the online version of the paper.

Received 10 April 2012; accepted 26 July 2013.

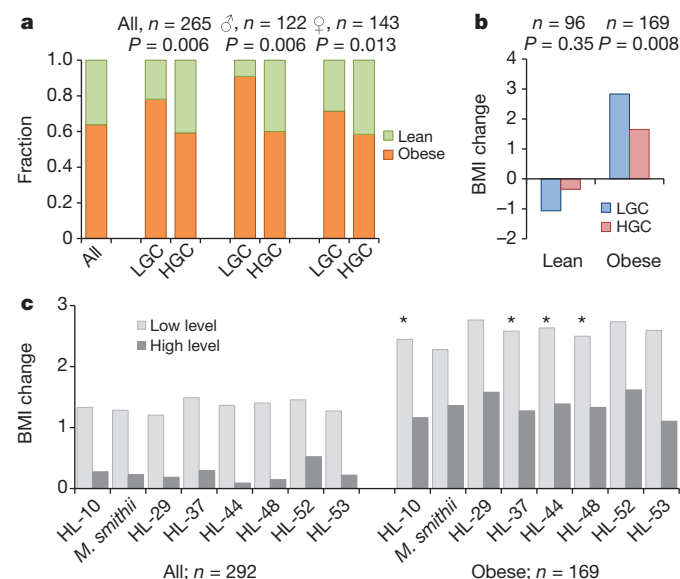


Figure 4 | Evolution of BMI in obese and non-obese LGC and HGC individuals (*n* = 265). **a**, LGC individuals are more frequently obese. **b**, LGC obese individuals gained more weight over 9 years. **c**, Bacterial species associated with weight change. BMI change was computed for at least 125 (all) or 60 (obese) individuals having undetectable or lowest level of a species ('Low level') and for at least 40 (all) or 30 (obese) individuals having the highest abundance of the same species ('High level'). Average BMI change of low and high level groups was significantly different (*P* < 0.05 or **P* < 0.01), with the exception of HL-52 for all individuals (*P* = 0.052); for obese, BMI and age were adjusted.

- World Health Organization. Obesity and overweight. Fact sheet no. 311; <http://www.who.int/mediacentre/factsheets/fs311/en/> (2006).
- Kelly, T., Yang, W., Chen, C. S., Reynolds, K. & He, J. Global burden of obesity in 2005 and projections to 2030. *Int. J. Obes.* **32**, 1431–1437 (2008).
- Stunkard, A. J., Harris, J. R., Pedersen, N. L. & McClearn, G. E. The body-mass index of twins who have been reared apart. *N. Engl. J. Med.* **322**, 1483–1487 (1990).
- Allison, D. B. *et al.* The heritability of body mass index among an international sample of monozygotic twins reared apart. *Int. J. Obes. Relat. Metab. Disord.* **20**, 501–506 (1996).
- Maes, H. H., Neale, M. C. & Eaves, L. J. Genetic and environmental factors in relative body weight and human adiposity. *Behav. Genet.* **27**, 325–351 (1997).

6. Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genet.* **42**, 937–948 (2010).
7. Frayling, T. M. *et al.* A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**, 889–894 (2007).
8. Loos, R. J. *et al.* Common variants near *MC4R* are associated with fat mass, weight and risk of obesity. *Nature Genet.* **40**, 768–775 (2008).
9. Willer, C. J. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nature Genet.* **41**, 25–34 (2009).
10. Thorleifsson, G. *et al.* Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nature Genet.* **41**, 18–24 (2009).
11. Heid, I. M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nature Genet.* **42**, 949–960 (2010).
12. Lindgren, C. M. *et al.* Genome-wide association scan meta-analysis identifies three loci influencing adiposity and fat distribution. *PLoS Genet.* **5**, e1000508 (2009).
13. Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: human gut microbes associated with obesity. *Nature* **444**, 1022–1023 (2006).
14. Furet, J. P. *et al.* Differential adaptation of human gut microbiota to bariatric surgery-induced weight loss: links with metabolic and low-grade inflammation markers. *Diabetes* **59**, 3049–3057 (2010).
15. Duncan, S. H. *et al.* Human colonic microbiota associated with diet, obesity and weight loss. *Int. J. Obes.* **32**, 1720–1724 (2008).
16. Schwirtz, A. *et al.* Microbiota and SCFA in lean and overweight healthy subjects. *Obesity* **18**, 190–195 (2010).
17. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457**, 480–484 (2009).
18. Backhed, F. *et al.* The gut microbiota as an environmental factor that regulates fat storage. *Proc. Natl Acad. Sci. USA* **101**, 15718–15723 (2004).
19. Ley, R. E. *et al.* Obesity alters gut microbial ecology. *Proc. Natl Acad. Sci. USA* **102**, 11070–11075 (2005).
20. Turnbaugh, P. J., Backhed, F., Fulton, L. & Gordon, J. I. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe* **3**, 213–223 (2008).
21. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).
22. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
23. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).
24. Cotillard, A. *et al.* Dietary intervention impact on gut microbial gene richness. *Nature* <http://dx.doi.org/10.1038/nature12480> (this issue).
25. Rajilic-Stojanovic, M. *et al.* Development and application of the human intestinal tract chip, a phylogenetic microarray: analysis of universally conserved phylotypes in the abundant microbiota of young and elderly adults. *Environ. Microbiol.* **11**, 1736–1751 (2009).
26. Manichanh, C. *et al.* Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* **55**, 205–211 (2006).
27. Lepage, P. *et al.* Twin study indicates loss of interaction between microbiota and mucosa of patients with ulcerative colitis. *Gastroenterology* **141**, 227–236 (2011).
28. Claesson, M. J. *et al.* Gut microbiota composition correlates with diet and health in the elderly. *Nature* **488**, 178–184 (2012).
29. Costello, E. K. *et al.* Bacterial community variation in human body habitats across space and time. *Science* **326**, 1694–1697 (2009).
30. Cho, I. *et al.* Antibiotics in early life alter the murine colonic microbiome and adiposity. *Nature* **488**, 621–626 (2012).
31. Vijay-Kumar, M. *et al.* Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science* **328**, 228–231 (2010).
32. Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227 (2012).
33. Sokol, H. *et al.* *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl Acad. Sci. USA* **105**, 16731–16736 (2008).
34. Devillard, E., McIntosh, F. M., Duncan, S. H. & Wallace, R. J. Metabolism of linoleic acid by human gut bacteria: different routes for biosynthesis of conjugated linoleic acid. *J. Bacteriol.* **189**, 2566–2570 (2007).
35. Png, C. W. *et al.* Mucolytic bacteria with increased prevalence in IBD mucosa augment *in vitro* utilization of mucin by other bacteria. *Am. J. Gastroenterol.* **105**, 2420–2428 (2010).
36. Swidsinski, A., Weber, J., Loening-Baucke, V., Hale, L. P. & Lochs, H. Spatial organization and composition of the mucosal flora in patients with inflammatory bowel disease. *J. Clin. Microbiol.* **43**, 3380–3389 (2005).
37. Joossens, M. *et al.* Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut* **60**, 631–637 (2011).
38. Yang, B. *et al.* Unsupervised binning of environmental genomic fragments based on an error robust selection of l-mers. *BMC Bioinformatics* **11** (suppl. 2), S5 (2010).
39. Qin, J. *et al.* A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55–60 (2012).
40. Chevalleyre, Y., Koriche, F. & Zucker, J.-D. Rounding methods for discrete linear classification. *Proc. 30th Int. Conf. Machine Learning (ICML-13)* 651–659 (2013).
41. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
42. Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105–108 (2011).
43. Ouchi, N., Parker, J. L., Lugus, J. J. & Walsh, K. Adipokines in inflammation and metabolic disease. *Nature Rev. Immunol.* **11**, 85–97 (2011).
44. Shoelson, S. E., Lee, J. & Goldfine, A. B. Inflammation and insulin resistance. *J. Clin. Invest.* **116**, 1793–1801 (2006).
45. Backhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. Host-bacterial mutualism in the human intestine. *Science* **307**, 1915–1920 (2005).
46. Backhed, F., Manchester, J. K., Semenkovich, C. F. & Gordon, J. I. Mechanisms underlying the resistance to diet-induced obesity in germ-free mice. *Proc. Natl Acad. Sci. USA* **104**, 979–984 (2007).
47. Mandard, S. *et al.* The fasting-induced adipose factor/angiopoietin-like protein 4 is physically associated with lipoproteins and governs plasma lipid levels and adiposity. *J. Biol. Chem.* **281**, 934–944 (2006).
48. Membrez, M. *et al.* Gut microbiota modulation with norfloxacin and ampicillin enhances glucose tolerance in mice. *FASEB J.* **22**, 2416–2426 (2008).
49. Ajslev, T. A., Andersen, C. S., Gamborg, M., Sorensen, T. I. & Jess, T. Childhood overweight after establishment of the gut microbiota: the role of delivery mode, pre-pregnancy weight and early administration of antibiotics. *Int. J. Obes.* **35**, 522–529 (2011).
50. Sun, Y. *et al.* Advanced computational algorithms for microbial community analysis using massive 16S rRNA sequence data. *Nucleic Acids Res.* **38**, e205 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors wish to thank A. Forman, T. Lorentzen, B. Andreassen, G. J. Klavens and M. M. Andersen for technical assistance; A. L. Nielsen, G. Lademann and M. M. H. Kristensen for management assistance; K. Kiil for discussions and assistance, and A. Walker for comments on the manuscript. This research has received funding from European Community's Seventh Framework Program (FP7/2007-2013): MetaHIT, grant agreement HEALTH-F4-2007-201052. Additional funding came from The Lundbeck Foundation Centre for Applied Medical Genomics in Personalized Disease Prediction, Prevention and Care (LuCamp, <http://www.lucamp.org>), ANR MicroObes, the Metagenopolis grant ANR-11-DPBS-0001, Region Ile de France (CODDIM) and Fondacoeur. The Novo Nordisk Foundation Center for Basic Metabolic Research is an independent Research Center at the University of Copenhagen partially funded by an unrestricted donation from the Novo Nordisk Foundation (<http://www.metabol.ku.dk>).

Author Contributions O.P. and S.D.E. designed the study, O.P., S.D.E., P.B., W.J., S.B., K.C., J.D., M.K., P.R., T.S.-P., W.M.d.V., T.H., J.R. and K.K. managed the study. T.N., K.B., T.H., N.G., T.J., I.B. and O.P. carried out patient phenotyping and clinical data analyses. T.N., K.B. and F.L. performed sample collection and DNA extraction. J.Q. and J.L. supervised DNA sequencing and gene profiling. S.D.E. and O.P. designed and supervised the data analyses. E.L.C., E.P., T.N., N.G., G.F., F.H., M.A.I., M.A.R., J.-M.B., S.K., P.L., N.P., S.S., J.T., J.Q., J.L., J.-D.Z., S.R. and S.D.E. performed the data analyses. S.T. and E.G.Z. carried out HITChip analysis. M.B., A.S.J., H.B.N. and T.S.-P. carried out metagenomic array analyses. S.D.E., O.P., J.R. and P.B. wrote the paper. MetaHIT consortium members provided creative environment and constructive criticism throughout the study.

Author Information The raw Illumina read data for all samples has been deposited in the EBI European Nucleotide Archive under the accession number ERP003612. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.B. (bork@embl.de, for bioinformatics), W.J. (wangj@genomics.org.cn, for sequencing), S.D.E. (dusko.ehrlich@jouy.inra.fr, for any aspect of this work) or O.P. (oluf@sund.ku.dk, for physiological aspects of this work).

MetaHIT consortium additional members

Eric Guedon¹, Christine Delorme¹, Séverine Layec¹, Ghali Khaci¹, Maarten van de Guchte¹, Gaetana Vandemeulebrouck¹, Alexandre Jamet¹, Rozenn Dervyn¹, Nicolas Sanchez¹, Emmanuelle Maguin¹, Florence Haimet², Yohanan Winogradski¹, Antonella Cultrone¹, Marion Leclerc¹, Catherine Juste¹, Hervé Blottière^{1,2}, Eric Pelletier^{3,4,5}, Denis LePaslier^{3,4,5}, François Artiguenave^{3,4,5}, Thomas Bruls^{3,4,5}, Jean Weissenbach^{3,4,5}, Keith Turner⁶, Julian Parkhill⁶, Maria Antolin⁷, Chaysavanh Manichanh⁷, Francesc Casellas⁷, Natalia Boruel⁷, Encarna Varela⁷, Antonio Torrejon⁷, Francisco Guarner⁷, Gérard Denariatz⁷, Muriel Derrien⁸, Johan E. T. van Hylckama Vlieg⁸, Patrick Veiga⁸, Raish Oozeer⁹, Jan Klot⁹, Maria Rescigno¹⁰, Christian Brechot¹¹, Christine M'Rini¹¹, Alexandre Mérieux¹¹ & Takuji Yamada¹²

¹INRA, Institut National de la Recherche Agronomique, UMR 14121 MICALIS, 78350 Jouy en Josas, France. ²INRA, Institut National de la Recherche Agronomique, US1367 Metagenopolis, 78350 Jouy en Josas, France. ³Commissariat à l'Energie Atomique, Genoscope, 91000 Evry, France. ⁴Centre National de la Recherche Scientifique, UMR8030, 91000 Evry, France. ⁵Evry, France, Université d'Evry Val d'Essonne. 91000 Evry, France. ⁶The Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK. ⁷Digestive System Research Unit, University Hospital Vall d'Hebron, Ciberehd, 08035 Barcelona, Spain. ⁸Danone Research, 91120 Palaiseau, France. ⁹Gut Biology & Microbiology, Danone Research, Centre for Specialized Nutrition, Bosrandweg 20, 6704 PH Wageningen, The Netherlands. ¹⁰Istituto Europeo di Oncologia, 20100 Milan, Italy. ¹¹Institut Mérieux, 17 rue Burgelat, 69002 Lyon, France. ¹²European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany.

METHODS

Study population. The study participants were recruited from the Inter99 study population. The Inter99 study is a randomized, non-pharmacological intervention study for the prevention of ischaemic heart disease, and was conducted at the Research Centre for Prevention and Health in Glostrup, Denmark, between 1999 and 2006 (clinicalTrials.gov: NCT00289237)⁵¹. The participants in the Inter99 study were examined at baseline, after 1, 3 and 5 years depending on the type of intervention.

For the present study individuals with BMI < 25 kg m⁻² or BMI > 30 kg m⁻² at year five in the Inter99 study were randomly selected from track records. They had no known gastrointestinal disease, no previously bariatric surgery, no medications known to affect the immune system, and no antibiotics 2 months before faecal sample collection. Individuals with type-2 diabetes at the day of examination were excluded. In total, 292 non-diabetic individuals were included in the protocol. All had North European ethnicity. At the time of the current physical examination, 96 (33%) of study volunteers were lean with BMI < 25 kg m⁻², 27 (9%) were overweight with BMI between 25 and 30 kg m⁻², and 169 (58%) were obese with BMI > 30 kg m⁻², according to World Health Organization definitions⁵². The study was approved by the local Ethical Committees of the Capital Region of Denmark (HC-2008-017), and was in accordance with the principals of the Declaration of Helsinki. All individuals gave written informed consent before participation in the study.

Phenotyping. The participants were examined on two different dates, approximately 14 days apart. On the first day, participants were examined in the morning after an overnight fast. Height was measured without shoes to the nearest 0.5 cm, and weight was measured without shoes and wearing light clothes to the nearest 0.1 kg. Hip and waist circumference were recorded using a non-expandable measuring tape to the nearest 0.5 cm. Waist circumference was measured midway between the lower rib margin and the iliac crest. Hip circumference was measured as the largest circumference between the waist and the thighs. On the second day of examination, all participants delivered a stool sample collected at home and dual-emission X-ray absorptiometry (DXA) was performed. Analyses of data from the DXA scan were conducted with the integrated software (Hologic Discovery A, Santax). Sagittal height was measured at the time of the DXA scan with the use of the Holtain-Kahn abdominal caliper at the highest point of the abdomen with the participant supine and while breathing out. Participant receiving statins, fibrates and/or ezetimibe were reported as receiving lipid-lowering medication.

Derived anthropometrical measure and index of insulin resistance. Intra-abdominal adipose tissue (cm²) was calculated using data from DXA scans and anthropometry using the equation⁵³: $y = -208.2 + 4.62 (\text{sagittal diameter, cm}) + 0.75 (\text{age, years}) + 1.73 (\text{waist, cm}) + 0.78 (\text{trunk fat, \%})$. HOMA-IR was calculated as: $(\text{fasting plasma glucose (mmol l}^{-1}) \times \text{fasting serum insulin (mU l}^{-1}) / 22.5)$ (ref. 54).

Biochemical measurements. All analyses were performed on blood samples drawn in the morning after an overnight fast from at least 22:00 the previous evening.

Plasma glucose was analysed by a glucose oxidase method (Granustest, Merck) with a detection limit of 0.11 mmol l⁻¹ and intra- and inter-assay coefficients of variation (CV) of <0.8 and <1.4%, respectively. HbA1c was measured on TOSOH G7 by ion-exchange high performance liquid chromatography.

Serum insulin (excluding des-31,32-proinsulin and intact proinsulin) was measured using the AutoDELFIA insulin kit (Perkin-Elmer, Wallac) with a detection limit of 3 pmol l⁻¹ and with intra- and inter-assay CV of <3.2% and <4.5%, respectively. Plasma total cholesterol, plasma HDL-cholesterol and plasma triglycerides were all measured on Vitros 5600 using reflect-spectrophotometrics. Blood leucocytes and white blood cell differential count were measured on Sysmex XS 1000i using flow cytometrics. Plasma ALT and plasma total free fatty acids were analysed using standard biochemical methods (Modular Evo). Plasma hsCRP was analysed by a particle-enhanced immunoturbidimetric assay on MODULAR Evo using CRPL3 kit (Roche) with a detection limit of 0.3 mg l⁻¹ and intra- and inter-assay CV of <4.0% and <6.2%, respectively.

Serum adiponectin was analysed using a two-site-sandwich ELISA kit for measuring total human adiponectin (TECO). Detection limit was 0.6 ng ml⁻¹ and intra- and inter-assay CV were <4.66% and <6.72%, respectively. Serum FIAF was measured using a quantitative sandwich ELISA (Adipo Bioscience). Detection limit was 0.6 µg l⁻¹ and the intra- and inter-assay CV of 4% and 8%, respectively. Serum lipopolysaccharide binding protein was analysed by a solid phase sandwich ELISA kit (Abnova) with intra- and inter-assay CV of <6.1% and <17.8%, respectively. Serum IL-6 and serum TNF-α were analysed by Luminex using the Bio-Plex Pro cytokine assay (Bio-Rad), whereas serum leptin was measured using the Bio-Plex Pro diabetes assay.

Faecal sampling. Stool samples were obtained at the homes of each participant and samples were immediately frozen in their home freezer. Frozen samples were delivered to Steno Diabetes Center using insulating polystyrene foam containers,

and stored at -80 °C until analysis. The time span from sampling to delivery at the Steno Diabetes Center was aimed to be as short as possible and no more than 48 h. **DNA extraction.** A frozen aliquot (200 mg) of each faecal sample was suspended in 250 µl of guanidine thiocyanate, 0.1 M Tris, pH 7.5, and 40 µl of 10% N-lauroyl sarcosine. Then, DNA extraction was conducted as previously described²⁶. The DNA concentration and its molecular size were estimated by nanodrop (Thermo Scientific) and on agarose gel electrophoresis.

Illumina sequencing. DNA library preparation followed the manufacturer's instruction (Illumina). We used the workflow indicated by the provider to perform cluster generation, template hybridization, isothermal amplification, linearization, blocking and denaturing and hybridization of the sequencing primers. The base-calling pipeline (version IlluminaPipeline-0.3) was used to process the raw fluorescent images and call sequences.

We constructed one library (clone insert size 200 base pairs (bp)) for each of the first batch of 15 samples; two libraries with different clone insert sizes (135 and 400 bp) for each of the second batch of 70 samples, and one library (350 bp) for each of the third batch of 207 samples.

After sequencing, we performed quality control and screened human genome contaminant. Finally, we generated 26.0 million–186.1 million high-quality reads for the 292 samples, with an average of 68.2 million high-quality reads. Sequencing read length of the first batch of 15 samples was 44 bp, the second batch was 75 bp, and the third batch was 75 bp and 90 bp.

Microbial gene abundance profiling by quantitative metagenomics. An average of 34.1 million paired-end reads were produced for each sample and, after removing human contamination (~0.1%, on average), 19.9 ± 6.7 million reads were mapped at a unique position of the reference catalogue of 3.3 million genes, using SOAP2.21 (ref. 55) by allowing at most two mismatches in the first 35-bp region and 90% identity over the read sequence; reads mapping at multiple positions (13.4%, on average) were discarded. The abundance of a gene in a sample was estimated by dividing the number of reads that uniquely mapped to that gene by the gene length and by the total number of reads from the sample that uniquely mapped to any gene in the catalogue. The resulting set of gene abundances, termed a microbial gene profile of an individual, was used for further analyses; Illumina and Solid sequencing platforms gave highly similar gene profiles for the same DNA samples²⁴.

Gene counting. On the basis of the pair-oriented counting result of each samples, we selected the threshold of one read for gene identification to include the rare genes into the analysis. We identified 91,032–1,005,488 genes for the 292 samples, with an average of 670,528 genes.

Read downsizing. To eliminate the influence of sequencing depth fluctuation, we sampled the alignment results and downsized the number of mapped pairs to 11 million for each sample. After that, we found 59,147–878,816 genes for the 292 samples, with an average of 578,512 genes.

Diversity estimate by single copy gene scoring. Genes belonging to the orthologous groups COG0085, COG052, and COG0090 from 3,515 prokaryotic genomes were clustered to OTUs at 95% identity by UCLUST and used as a reference database. Paired-end Illumina reads from 292 metagenomic samples were mapped at 95% identity cut-off using soap2.21 (ref. 55). The numbers of fragments that were assigned to the reference sequences were counted so that each fragment's weight equals 1, that is, a fragment assigned to *N* different reference sequences contributes 1/*N* to each reference sequence. Fragment counts of reference sequences were grouped to yield OTU counts. Samples with low sampling effort, that is, with less than 3,000 fragments mapped to reference genes, were removed leaving 229 samples for comparative analyses. OTU counts were normalized by gene length, scaled by the maximum count across all marker genes, and down-sampled using the vegan package to the minimum sum of OTU counts across all samples to compare species richness between high and low gene content groups.

Phylogenetic microarray analysis. HITChip microarray analyses were performed as described previously²⁵. In short, 16S rRNA genes were amplified the T7prom-Bact-27-for and Uni-1492-rev primers from 10 ng fecal DNA extracts. On these amplicons an *in vitro* transcription and subsequent labelling with Cy3 and Cy5 dyes were performed. Labelled RNA was fragmented and hybridized on the arrays at 62.5 °C for 16 h in a rotation oven (Agilent Technologies). The arrays were washed, dried, scanned and the signal intensity data was extracted as described (<http://www.agilent.com>). Microarray data normalization and analysis were carried out with a set of R-based scripts (<http://r-project.org>), while making use of a custom designed database, which operates under the MySQL database management system (<http://www.mysql.com>).

From the 3,699 unique HITChip probes, we selected the probes that accounted for the top 99.9% of the total signal. These probes were counted for each sample to measure richness, which was between 713 and 1,597 probes per sample. The probes that accounted for the lowest 0.1% of the total signal were regarded as background

noise and were not taken into account for further analysis. Probe signal values were used to calculate the inverse Simpson's diversity index for each sample.

HITChip probes specificity can be assigned to three phylogenetic levels based on 16S rRNA gene sequence similarity: order-like groups, genus-like groups (sequence similarity >90%), and phylotype-like groups (sequence similarity >98%)²⁴. Relative abundances were calculated for each specificity level by summing all signal values of the probes targeting a group and dividing by the total of all probe signals for the corresponding sample. All comparisons between the HGC and LGC individuals were assessed with dependent two-group Wilcoxon signed rank tests. When statistical tests were performed on a large number of variables the obtained *P* values were adjusted by a Bonferroni correction. To place the gene count and BMI marker species (HL and ob/ob, referring to species significantly different between HGC/LGC and obese/lean individuals, respectively) in HITChip phylogeny, Spearman correlation coefficients were calculated between the metagenomic profiling frequencies and relative abundances of the phylotype-like across 251 samples. A threshold of 0.7 was used to associate 16S to a species.

The Pearson correlation coefficient between the log(gene count) and log(probe count) was 0.8 and the concordance between assignments of individuals to a low or high richness class by the two techniques (gene counts or probe counts) was 88%. **Metagenomic microarray analysis.** A 2.1-million-feature custom Roche NimbleGen microarray targeting a 700,000 genes subset of the MetaHit human gut gene catalogue²² was designed and manufactured. The subset of genes was prioritized for genes that were observed in more than 20 of the 124 gene catalogue samples. DNA extracted from faecal samples were labelled and hybridized according to standard NimbleGen protocols. Data was pre-processed and Shannon diversity index calculated using the RMA implementation under the 'oligo' package and the vegan package, respectively, both available in the statistical programming environment R.

To validate the observed biomarkers for low/high gene counts found by sequencing, we compared the data to DNA microarray signals for the same samples and individuals. Thus, the tracer genes for known and unknown species indicated in Fig. 2 were compared to a microarray gene set comprising more than 700,000 gut-associated genes selected from the MetaHit Gene Catalog⁸ in addition to reference genomes. Perfect matches were found for 129 tracer genes on the DNA microarray. To test whether a similar discrimination could be obtained from the microarray data, the samples were divided into low and high diversity sets using the Shannon diversity index. Using this index, 90 samples were categorized as low diversity, and 70 were categorized as high. Differences in DNA abundance signals between low and high diversity samples were tested for the 129 matching genes (*t*-test). Summarized, in terms of species the following groups were associated to low diversity, *C. clostridioforme*/C. bolteae, HL-7, HL-39, *R. gnavus*, HL-15 and *Bacteroides*, whereas HL-53 and *M. smithii* were associated to high diversity. These DNA microarray observations support the quantitative metagenomics results (Fig. 2 and Supplementary Table 4).

Phylogenetic annotation. Taxonomic assignment of predicted genes for global analysis was carried out using BLASTN to assign reads to a reference genome database at a cut-off of 95% sequence identity and >100 bp overlap, unless indicated otherwise. This assignment was used as high confidence assignment on species level. As reference database we used 1,869 available reference genomes from NCBI and the set of draft gastrointestinal genomes from the DACC (<http://hmpdacc.org/>), both as of the 15 July 2011 version. The assigned reads to each taxonomic group per sample were rarefied to 5.5 million reads (the size of the smallest sample), on this rarefied matrix taxonomic groups were tested for significant differences in abundance using a Wilcoxon rank-sum test. Multiple testing corrections were performed using the Benjamini–Hochberg method⁴⁰ ($q < 0.1$). From the same matrix we calculated the genus abundance as the percentage of reads assigned from 5.5 million total reads per sample; this matrix was used to calculate the class-wise means as well as standard deviations expressed in percentages (Supplementary Table 2).

Functional annotation. BLASTP was used to search the protein sequences of the predicted genes in the eggNOG database⁵⁶ and KEGG database⁵⁷ with $E \leq 10^{-5}$ as described previously⁸, and the NOG/KEGG OG of the best hit was assigned to each gene. The genes annotated by COG were classified into the 25 COG categories, and genes that were annotated by KEGG were assigned to a set of manually determined gut metabolic modules (G.W. *et al.*, manuscript in preparation). The relative pathway/module abundance of higher order functional categories were calculated from rarefied KO abundances. Modules were deemed present when $\geq 30\%$ of the enzymes were recovered, after manual removing of overly 'promiscuous' enzymes (that is, present in multiple modules) before abundance calculation. For higher-level functional assignments, KO abundances were summed and distributed evenly when KO groups appeared in multiple categories. Functional differences were calculated with a Wilcoxon rank-sum test and multiple testing corrections were performed using the Benjamini–Hochberg method⁴⁰ ($q < 0.05$).

Genes significantly different in groups of individuals. Genes significantly different in abundance between groups of individuals were identified by the Wilcoxon rank-sum test coupled to a bootstrapping approach.

Approximately 70% of the whole cohort (204 individuals) was randomly chosen and genes differentially abundant between LGC and HGC individuals were identified at $P \leq 0.0001$ as threshold. This test was repeated 30 times. We also composed 30 groups of randomly chosen 'extreme' individuals that had <400,000 or >600,000 genes and applied the same test. Genes common to all 60 tests were further analysed.

For lean and obese individuals we used a similar approach by randomly choosing 70% of individuals 30 times and using Wilcoxon rank-sum test at $P \leq 0.05$.

Gene clustering method. We used an unsupervised strategy to cluster genes of the same species. Such genes are expected to be present at a similar abundance in an individual but at different abundances in different individuals. The genes that vary in abundance in a coordinated way are thus likely to be from the same species. The clustering algorithm developed in Delphi6 programming language consists of two steps: (1) Spearman correlation coefficients were determined for all pairs of genes of each gene set, using the covariance abundance of the genes among the 292 individuals, and the SpearmanRankCorrelation function included in the free delphi correlation.pas library of the ALGLIB project (see <http://www.alglib.net/>). (2) All the genes correlated above a given threshold were assigned to the same cluster. If a gene of a given cluster was correlated with a gene of another cluster above the threshold, the genes of the two clusters were merged (single linkage algorithm).

A similar clustering approach was recently described, using Kendall tau instead of Spearman R to compute correlation coefficients and double linkage (all genes of a cluster are correlated above the threshold) rather than single linkage³⁸. We have not compared the performance of the two methods, but rather characterized the outcome of the one we used.

Cluster characterization. To test whether the tracer genes originate from the same species we carried out a number of analyses, using as control 135 clusters composed of randomly chosen catalogue genes.

(1) BLASTN. Some 17 clusters contained genes that matched a reference genome at a threshold of 95% identity over 90% of gene length when mapped against a collection of 6,006 genomes (the available reference genomes from NCBI and the set of draft gastrointestinal genomes from the DACC and MetaHIT as of the 3 August 2012 version). The taxonomic assignment was highly coherent (Supplementary Fig. 5). We denote these clusters hereafter as 'known', referring to their taxonomy at the above-mentioned thresholds. Similarly, for clusters containing genes that did not reach these thresholds, the genes were uniformly not assigned to known genomes. We denote these clusters as 'unknown'. A single exception to the rule was a cluster having a high identity (97–99%, on average) with four different genomes, *Pseudoflavonifractor capilosus*, *F. prausnitzii* L2-6, *F. prausnitzii* A2-16 and *Subdoligranulum variabile* DSM16176. The genes of this cluster were present in the same order on all genomes (Supplementary Fig. 9); we suggest that they derive from a (possibly defective) prophage or a conjugative transposon integrated in the chromosome (BLASTP analysis reveals functions involved in DNA metabolism and transfer, such as DNA helicases or tra genes). A similar grouped localization of the cluster genes was observed for the *Roseburia inulinivorans* genome revealed in the accompanying manuscript (cluster MO-HL-16)²⁴, suggesting that they also originate from a prophage or a transposon. These observations point to the capacity of our clustering method to group genes not only of bacteria but also of mobile and promiscuous elements, and indicate a potential source of false positives, as regards grouping of genes from a single species. Identification of all the catalogue genes that co-vary by abundance with the tracer genes could probably help to differentiate clusters that represent bacterial species from those that originate from promiscuous genetic elements, as the former should include substantially more genes than the latter (species encode mostly >500 genes while the promiscuous genetic elements encode generally <100). Such systematic analysis is, however beyond the scope of the current work.

(2) Abundance covariance of the genes of each cluster was computed across all individuals of our cohort and also across the French cohort from the accompanying manuscript²⁴. The genes of both known and unknown clusters co-varied by abundance in the two data sets (Danish and French, Supplementary Fig. 6; the values were higher and more homogeneous in the former than the latter, presumably reflecting the data set size: 292 Danes and 49 French). For the randomly constituted groups pairwise correlation values were close to zero (0.01 ± 0.006).

(3) Abundance similarity. Abundance covariance does not imply that the genes of a cluster have a similar abundance in an individual, as only the abundance ratios could be constant. Nevertheless it is expected that the genes belonging to the same species show similar abundance profiles as they are carried by the same DNA molecule; this is a basic postulate of our approach to gene clustering. To estimate abundance similarity we computed for each cluster a score of gene

abundance variation for each individual as a ratio of the standard deviation of the gene abundance over the mean in each individual. The known and unknown clusters show very small and similar scores indicating the homogeneity of the abundance signal. That is not the case for the randomly constituted groups (Supplementary Fig. 7).

(4) *k*-mer composition. This composition is similar across a given genome and different from that of unrelated genomes³⁷. We computed first the tetranucleotide profile of each gene of a cluster and then the pairwise correlation of the profiles of all genes of a cluster. Genes that belong to the known and unknown clusters have similar highly correlated 4-mer profiles, whereas the genes of the control random groups have a much lower correlation (Supplementary Fig. 8, left). Because the cluster gene size ($1,135 \pm 446$ bp) is relatively low, we retrieved from the catalogue the contigs that carry the genes of each group. Indeed, about 2,000 bp is estimated to be necessary for accurate measure of the 135 non-redundant 4-mer combinations³⁷. The contigs are longer than genes ($4,084 \pm 3,678$ bp) and have a higher correlation of 4-mer profiles (Supplementary Fig. 8, right).

(5) Physical linkage of the genes. We retrieved from the gene catalogue the scaffolds (composed of contigs that are 'bridged' by a paired-end sequence; one end is in one contig and the other in another, but the two sequences are not contiguous) that bear the tracer genes and searched for the presence of more than one gene of a cluster on a scaffold. Most of the known and unknown groups had at least some genes on the same scaffold (on average, about 18%) whereas randomly composed clusters had none, a difference of high significance ($P < 2 \times 10^{-16}$).

We conclude that the genes from a cluster originate, as expected, from the same genome.

As only a minority of the tracer gene clusters (15%) could be assigned to species-level taxonomy by BLASTN, we used BLASTP against either a collection of 6,006 available genomes or the non-redundant sequences databases available at NCBI to assess the taxonomy of other clusters. On the basis of the criterion of the homogeneity of the best-hit taxonomic assignment (with the threshold of at least 80% of a cluster genes having the same taxonomic best hit assignment), 91.9% and 45.2% of the clusters could be assigned at a phylum and genus level, respectively (Supplementary Fig. 11 and Supplementary Tables 4 and 11). In all cases the higher-level taxonomic assignments were congruent with the lower level ones.

Species abundance determination. We assessed whether the genes of a cluster can be used as tracers for a species they originate from, using the known clusters as benchmark. We examined whether the genes are either homologous to only the cognate genome of a species when a single strain of a species was sequenced, or homologous to all cognate genomes when multiple strains of a species were sequenced. The results are summarized in Supplementary Table 5. Eight clusters were in the first category and five in the second. Three of the latter matched two genomes of inconsistent species annotation, but which belong to the same species as deduced from the BLASTN comparison of the two genomes, which had an average identity of the reciprocal best hit genes comprised between 99.5% and 99.8%. A single cluster matched two of the five sequenced *F. prausnitzii* genomes with 98% average identity and three other genomes with an identity of 80–84%. However, the three last genomes had an average reciprocal best-hit gene identity with the two well-matched genomes of only 80.4–86% and therefore should probably be considered as belonging to a different species. We conclude that, for the known groups, the cluster genes are present (1) on cognate genomes only, and (2) on all cognate genomes. By extrapolation, we suggest that the same holds true for unknown groups and that the cluster genes can be used as tracers for the species they derive from.

Abundance of a given species in each individual was estimated as a mean abundance of 50 tracer genes of each cluster. The values were very close to the mean frequency of all the genes of a cluster.

ROC analysis. The analyses were carried out to distinguish between HGC and LGC individuals or lean and obese individuals by a combination of bacterial species. For each combination, only a single decision model was considered. In this very specific regression model weights are only allowed to take the values in (0, −1, 1). More precisely, the weight of each species in a given combination that

belong to the set of the species more frequent in one group is equal to 1, whereas that of the species that belong to the set of species more frequent in the other group is equal to −1. The weight of each species that is outside of the combination is 0. For each individual, this model yields a DBA score. As opposed to the infinite number of regression models, such ternary models are finite and can be exhaustively explored. To select the best models, we used the cross-validated AUC criterion³⁹ well adapted to classification models for binary outcome data.

Species correlated with the BMI change. For the entire cohort of 292 individuals, 40 individuals (14%) having the highest abundance of a species were compared with at least 125 individuals (42%) having the lowest abundance (all individuals lacking a species were included, when more numerous than 125); these numbers were chosen to allow contrasting the extremes of the distribution while keeping the sample size high enough to reduce the probability of a fortuitous difference in BMI change. For the 169 obese individuals, 30 (18%) having the highest abundance of a species were compared with at least 60 individuals (36%) having the lowest abundance (all individuals lacking a species were included, when more numerous than 60). The differences were calculated with a Student's *t*-test, the BMI changes being normally distributed, and multiple testing correction was performed using the Benjamini–Hochberg method⁴¹ ($q < 0.05$).

Association of microbial composition and metabolic traits. We analysed the association of (1) the high gene and low gene group, and (2) gene count as a continuous trait to quantitative traits applying a linear model adjusting for age and sex. Plasma triglycerides, plasma HDL cholesterol, serum insulin, plasma ALT, serum leptin and serum adiponectin and HOMA-IR were log transformed, whereas blood leucocytes, lymphocytes, monocytes, neutrophilocytes, plasma hsCRP, serum FIAF, plasma free fatty acids, serum TNF- α , serum interleukin (IL)-6, serum lipopolysaccharide binding protein and BMI were rank normalized before analyses in the linear model. In the analyses of triglycerides, treatment with lipid lowering medications was added as a covariate to the linear model. We corrected for multiple testing by the Benjamini–Hochberg method⁴¹, setting the FDR at 10%.

Bacterial species that discriminate between lean and obese individuals. To assess the difference in bacterial species between the lean (BMI $< 25 \text{ kg m}^{-2}$, $n = 96$) and obese (BMI $> 30 \text{ kg m}^{-2}$, $n = 169$) individuals we searched for the genes significantly different by abundance. Some 15,894 were found at $P < 0.05$, a value lower than that for the LGC and HGC individuals, indicating that the gut microbiota of lean and obese individuals differs less. In total, 14,149 could be clustered into 187 groups by covariance, at a Spearman threshold of $\rho > 0.75$, and 90% of these (12,753) were found in only 18 groups, ranging from 2,507 to 68 genes (Supplementary Table 10); four were correlated with a 16S rRNA gene by a co-variance-based HITChip analysis (Supplementary Table 11). The species represented by the 18 clusters had a significantly different distribution among the lean and obese individuals (Supplementary Fig. 12 and Supplementary Table 10); 4 were more frequent among the obese and 14 among the lean individuals.

51. Jørgensen, T. *et al.* A randomized non-pharmacological intervention study for prevention of ischaemic heart disease: baseline results Inter99. *Eur. J. Cardiovasc. Prev. Rehabil.* **10**, 377–386 (2003).
52. World Health Organization. Preventing and managing the globalepidemic. Report of a WHO consultation. *World Health Organ. Tech. Rep. Ser.* **894**, 1–253 (2000).
53. Treuth, M. S., Hunter, G. R. & Kekes-Szabo, T. Estimating intraabdominal adipose tissue in women by dual-energy X-ray absorptiometry. *Am. J. Clin. Nutr.* **62**, 527–532 (1995).
54. Matthews, D. R. *et al.* Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* **28**, 412–419 (1985).
55. Li, R. *et al.* SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* **25**, 1966–1967 (2009).
56. Jensen, L. J. *et al.* eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic Acids Res.* **36**, D250–D254 (2008).
57. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40**, D109–D114 (2012).

A ‘kilonova’ associated with the short-duration γ -ray burst GRB 130603B

N. R. Tanvir¹, A. J. Levan², A. S. Fruchter³, J. Hjorth⁴, R. A. Hounsell³, K. Wiersema¹ & R. L. Tunnicliffe²

Short-duration γ -ray bursts are intense flashes of cosmic γ -rays, lasting less than about two seconds, whose origin is unclear^{1,2}. The favoured hypothesis is that they are produced by a relativistic jet created by the merger of two compact stellar objects (specifically two neutron stars or a neutron star and a black hole). This is supported by indirect evidence such as the properties of their host galaxies³, but unambiguous confirmation of the model is still lacking. Mergers of this kind are also expected to create significant quantities of neutron-rich radioactive species^{4,5}, whose decay should result in a faint transient, known as a ‘kilonova’, in the days following the burst^{6–8}. Indeed, it is speculated that this mechanism may be the predominant source of stable r-process elements in the Universe^{5,9}. Recent calculations suggest that much of the kilonova energy should appear in the near-infrared spectral range, because of the high optical opacity created by these heavy r-process elements^{10–13}. Here we report optical and near-infrared observations that provide strong evidence for such an event accompanying the short-duration γ -ray burst GRB 130603B. If this, the simplest interpretation of the data, is correct, then it confirms that compact-object mergers are the progenitors of short-duration γ -ray bursts and the sites of significant production of r-process elements. It also suggests that kilonovae offer an alternative, unbeamed electromagnetic signature of the most promising sources for direct detection of gravitational waves.

Short-duration γ -ray bursts (SGRBs) have long been recognized as a distinct subpopulation of γ -ray bursts¹⁴. If they are indeed produced by

compact binary mergers, SGRBs may provide a bright electromagnetic signal accompanying events detected by the next generation of gravitational-wave interferometers¹⁵. Localizing electromagnetic counterparts is an essential prerequisite to obtaining direct redshift measurements and to constraining further the astrophysics of the sources. However, the evidence supporting this progenitor hypothesis is essentially circumstantial: principally, many SGRBs seem to reside in host galaxies, or regions within their hosts, that lack ongoing star formation, which makes an origin in massive stars unlikely (in contrast to long-duration γ -ray bursts, which result from the core collapse of some short-lived massive stars¹⁶). Progress in studying SGRBs has been slow; NASA’s Swift satellite localizes only a handful per year, and they are typically faint, with no optical afterglow or unambiguous host galaxy found in some cases despite rapid searches with large (8-m class) telescopes.

GRB 130603B was detected by Swift’s Burst Alert Telescope on 2013 June 3 at 15:49:14 UT¹⁷, and its duration was measured to be $T_{90} \approx 0.18 \pm 0.02$ s in the 15–350-keV band¹⁸. The burst was also detected independently by the Konus instrument on NASA’s Wind spacecraft, which found a somewhat shorter duration, $T_{90} \approx 0.09$ s in the 18–1,160-keV band¹⁹. This places the burst unambiguously in the short-duration class, which is also supported by the absence of the bright supernova emission generally found to accompany low-redshift ($z \lesssim 0.5$), long-duration bursts (see below). The optical afterglow was discovered at the William Herschel Telescope²⁰ and found to overlie a galaxy previously detected in the Sloan Digital Sky Survey imaging of

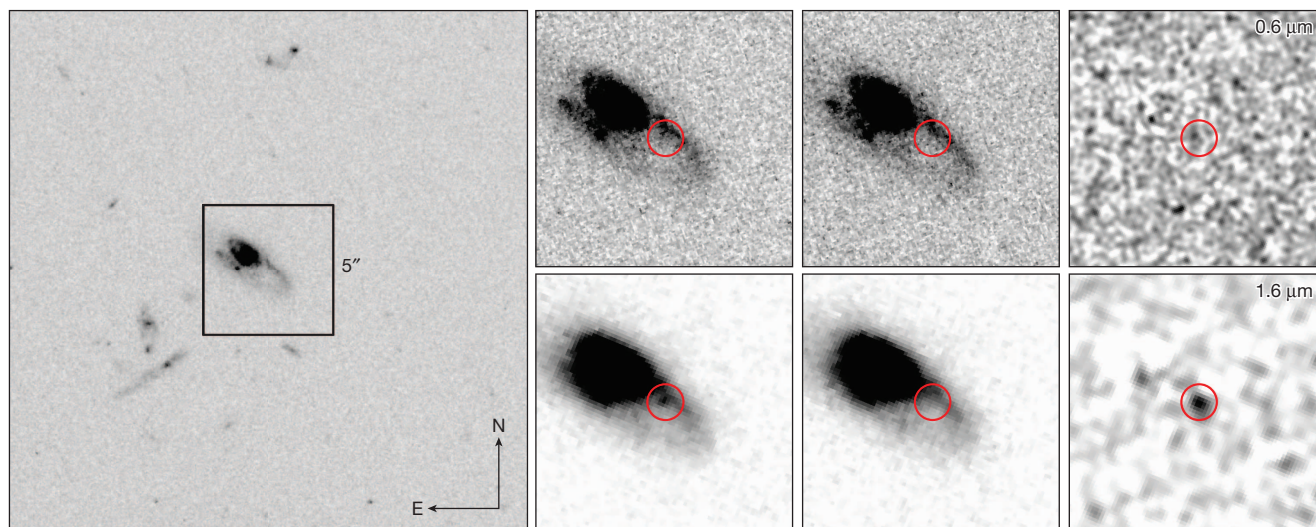


Figure 1 | HST imaging of the location of GRB 130603B. The host is well resolved and has a disturbed, late-type morphology. The position (coordinates $RA_{J2000} = 11$ h 28 min 48.16 s, $dec_{J2000} = +17^{\circ} 04' 18.2''$) at which the SGRB occurred (determined from ground-based imaging) is marked as a red circle (right-hand panels), lying slightly off a tidally distorted spiral arm. The left-hand panel shows the host and surrounding field from the higher-resolution optical image. The right-hand panels show, from left to right, the epoch-1 and epoch-2

imaging and their difference (epoch 1 minus epoch 2; upper row, F606W/ optical; lower row, F160W/NIR). The difference images have been smoothed with a Gaussian of width similar to the point-spread function, to enhance any point-source emission. Although the resolution of the NIR image is inferior to that of the optical image, we clearly detect a transient point source that is absent in the optical.

¹Department of Physics and Astronomy, University of Leicester, University Road, Leicester LE1 7RH, UK. ²Department of Physics, University of Warwick, Coventry CV4 7AL, UK. ³Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA. ⁴Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, 2100 Copenhagen, Denmark.

this field. The redshifts of the afterglow²¹ and the host galaxy²² were both found to be $z = 0.356$.

Another proposed signature of the merger of two neutron stars or a neutron star and a black hole is the production of a kilonova (sometimes also termed a ‘macronova’ or an ‘r-process supernova’) due to the decay of radioactive species produced and initially ejected during the merger process—in other words, an event similar to a faint, short-lived supernova^{6–8}. Detailed calculations suggest that the spectra of such kilonova sources will be determined by the heavy r-process ions created in the neutron-rich material. Although these models^{10–13} are still far from being fully realistic, a robust conclusion is that the optical flux will be greatly diminished by line blanketing in the rapidly expanding ejecta, with the radiation emerging instead in the near-infrared (NIR) and being produced over a longer timescale than would otherwise be the case. This makes previous limits on early optical kilonova emission unsurprising²³. Specifically, the NIR light curves are expected to have a broad peak, rising after a few days and lasting a week or more in the rest frame. The relatively modest redshift and intensive study of GRB 130603B made it a prime candidate for searching for such a kilonova.

We imaged the location of the burst with the NASA/ESA Hubble Space Telescope (HST) at two epochs, the first ~ 9 d after the burst (epoch 1) and the second ~ 30 d after the burst (epoch 2). On each occasion, a single orbit integration was obtained in both the optical F606W filter (0.6 μm) and the NIR F160W filter (1.6 μm) (full details of the imaging and photometric analysis discussed here are given in Supplementary Information). The HST images are shown in Fig. 1; the key result is seen in the difference frames (right-hand panels), which provide clear evidence for a compact transient source in the NIR in epoch 1 (we note that this source was also identified²⁴ as a candidate kilonova in independent analysis of our data on epoch 1) that seems to have disappeared by epoch 2 and is absent to the depth of the data in the optical.

At the position of the SGRB in the difference images, our photometric analysis gives a magnitude limit in the F606W filter of $R_{606,AB} > 28.25$ mag (2σ upper limit) and a magnitude in the F160W filter of $H_{160,AB} = 25.73 \pm 0.20$ mag. In both cases, we fitted a model point-spread function and estimated the errors from the variance of the flux at a large number of locations chosen to have a similar background to that at the position of the SGRB. We note that some transient emission may remain in the second NIR epoch; experimenting with adding synthetic stars to the image leads us to conclude that any such late-time emission is likely to be less than $\sim 25\%$ of the level in epoch 1 if it is not to appear visually as a faint point source in epoch 2, however, that would still allow the NIR magnitude in epoch 1 to be up to ~ 0.3 mag brighter.

To assess the significance of this result, it is important to establish whether any emission seen in the first HST epoch could have a contribution from the SGRB afterglow. A compilation of optical and NIR photometry, gathered by a variety of ground-based telescopes in the few days following the burst, is plotted in Fig. 2 along with our HST results. Although initially bright, the optical afterglow light curve declines steeply after about ~ 10 h, requiring a late-time power-law decay rate of $\alpha \approx 2.7$ (where $F \propto t^{-\alpha}$ describes the flux). The NIR flux, on the other hand, is significantly in excess of the same extrapolated power law. This point is made most forcibly by considering the colour evolution of the transient, defined as the difference between the magnitudes in each filter, which evolves from $R_{606} - H_{160} \approx 1.7 \pm 0.15$ mag at about 14 h to greater than $R_{606} - H_{160} \approx 2.5$ mag at about 9 d. It would be very unusual, and in conflict with predictions of the standard external-shock theory²⁵, for such a large colour change to be a consequence of late-time afterglow behaviour. The most natural explanation is therefore that the HST transient source is largely due to kilonova emission, and the brightness is in fact well within the range of recent models plotted in Fig. 2, thus supporting the proposition that kilonovae are likely to be important sites of r-process element production. We note that this phenomenon is strikingly reminiscent, in a qualitative sense, of the humps in the optical light curves of long-duration γ -ray bursts

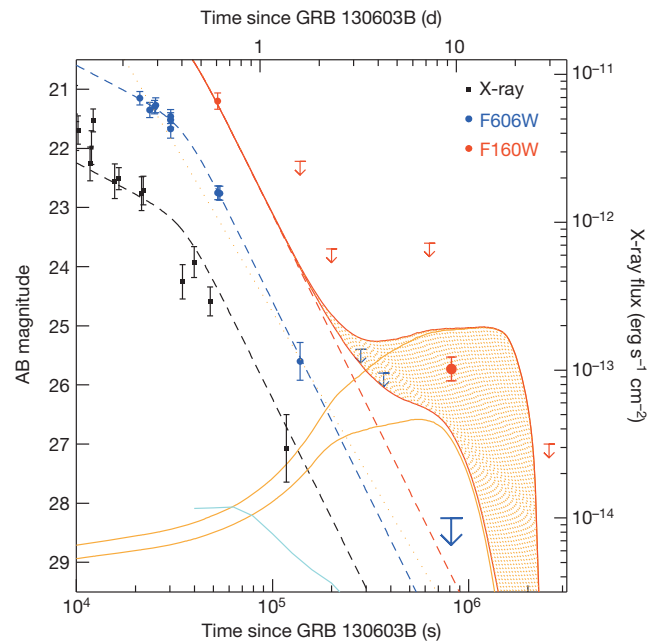


Figure 2 | Optical, NIR and X-ray light curves of GRB 130603B. Left axis, optical and NIR; right axis, X-ray. Upper limits are 2σ and error bars are 1σ . The optical data (g, r and i bands) have been interpolated to the F606W band and the NIR data have been interpolated to the F160W band using an average spectral energy distribution at ~ 0.6 d (Supplementary Information). HST epoch-1 points are given by bold symbols. The optical afterglow decays steeply after the first ~ 0.3 d and is modelled here as a smoothly broken power law (dashed blue line). We note that the complete absence of late-time optical emission also places a limit on any separate ^{56}Ni -driven decay component. The 0.3–10-keV X-ray data²⁹ are also consistent with breaking to a similarly steep decay (the dashed black line shows the optical light curve simply rescaled to match the X-ray points in this time frame), although the source had dropped below Swift sensitivity by ~ 48 h after the burst. The key conclusion from this plot is that the source seen in the NIR requires an additional component above the extrapolation of the afterglow (red dashed line), assuming that it also decays at the same rate. This excess NIR flux corresponds to a source with absolute magnitude $M(J)_{AB} \approx -15.35$ mag at ~ 7 d after the burst in the rest frame. This is consistent with the favoured range of kilonova behaviour from recent calculations (despite their known significant uncertainties^{11–13}), as illustrated by the model¹¹ lines (orange curves correspond to ejected masses of 10^{-2} solar masses (lower curve) and 10^{-1} solar masses (upper curve), and these are added to the afterglow decay curves to produce predictions for the total NIR emission, shown as solid red curves). The cyan curve shows that even the brightest predicted r-process kilonova optical emission is negligible.

produced by underlying type Ic supernovae, although here the luminosity is considerably fainter and the emission is redder. The ubiquity and range of properties of the late-time red transient emission in SGRBs will undoubtedly be tested by future observations.

The next generation of gravitational-wave detectors (Advanced LIGO and Advanced VIRGO) is expected ultimately to reach sensitivity levels allowing them to detect neutron-star/neutron-star and neutron-star/black-hole inspirals out to distances of a few hundred megaparsecs²⁶ ($z \approx 0.05$ – 0.1). However, no SGRB has been definitively found at any redshift less than $z = 0.12$ over the 8.5 yr of the Swift mission to date²⁷. This suggests either that the rate of compact binary mergers is low, implying a correspondingly low expected rate of gravitational-wave transient detections, or that most such mergers are not observed as bright SGRBs. The latter case could be understood if the beaming of SGRBs was rather narrow, for example, and the intrinsic event rate was, as a result, two or three orders of magnitude higher than that observed by Swift. Although the evidence constraining SGRB jet opening angles is limited at present²⁸ (indeed, the light-curve break seen in GRB 130603B may be further evidence for such beaming), it is clear that an alternative electromagnetic signature, particularly if approximately isotropic,

such as kilonova emission, could be highly important in searching for gravitational-wave transient counterparts.

Received 16 July; accepted 26 July 2013.

Published online 3 August 2013.

1. Nakar, E. Short-hard gamma-ray bursts. *Phys. Rep.* **442**, 166–236 (2007).
2. Lee, W. H. & Ramirez-Ruiz, E. The progenitors of short gamma-ray bursts. *New J. Phys.* **9**, 17–92 (2007).
3. Fong, W. *et al.* Demographics of the galaxies hosting short-duration gamma-ray bursts. *Astrophys. J.* **769**, 56–73 (2013).
4. Rosswog, S. *et al.* Mass ejection in neutron star mergers. *Astron. Astrophys.* **341**, 499–526 (1999).
5. Freiburghaus, C., Rosswog, S. & Thielemann, F.-K. R-process in neutron star mergers. *Astrophys. J.* **525**, L121–L124 (1999).
6. Li, L.-X. & Paczynski, B. Transient events from neutron star mergers. *Astrophys. J.* **507**, L59–L62 (1998).
7. Rosswog, S. Mergers of neutron star-black hole binaries with small mass ratios: nucleosynthesis, gamma-ray bursts, and electromagnetic transients. *Astrophys. J.* **634**, 1202–1213 (2005).
8. Metzger, B. D. *et al.* Electromagnetic counterparts of compact object mergers powered by the radioactive decay of r-process nuclei. *Mon. Not. R. Astron. Soc.* **406**, 2650–2662 (2010).
9. Gorieli, S., Bauswein, A. & Janka, H.-T. R-process nucleosynthesis in dynamically ejected matter of neutron star mergers. *Astrophys. J.* **738**, L32–L37 (2011).
10. Kasen, D., Badnell, N. R. & Barnes, J. Opacities and spectra of the r-process ejecta from neutron star mergers. *Astrophys. J.* (submitted); preprint at <http://arxiv.org/abs/1303.5788> (2013).
11. Barnes, J. & Kasen, D. Effect of a high opacity on the light curves of radioactively powered transients from compact object mergers. *Astrophys. J.* (submitted); preprint at <http://arxiv.org/abs/1303.5787> (2013).
12. Tanaka, M. & Hotokezaka, K. Radiative transfer simulations for neutron star merger ejecta. *Astrophys. J.* (in the press); preprint at <http://arxiv.org/abs/1306.3742> (2013).
13. Grossman, D., Korobkin, O., Rosswog, S. & Piran, T. The longterm evolution of neutron star merger remnants II: radioactively powered transients. *Mon. Not. R. Astron. Soc.* (submitted); preprint at <http://arxiv.org/abs/1307.2943> (2013).
14. Kouveliotou, C. *et al.* Identification of two classes of gamma-ray bursts. *Astrophys. J.* **413**, L101–L104 (1993).
15. Abadie, J. *et al.* Predictions for the rates of compact binary coalescences observable by ground-based gravitational-wave detectors. *Class. Quantum Gravity* **27**, 173001 (2010).
16. Hjorth, J. The supernova-gamma-ray burst-jet connection. *Phil. Trans. R. Soc. A* **371**, 20120275 (2013).
17. Melandri, A. *et al.* GRB 130603B: Swift detection of a bright short burst. *GCN Circ.* **14735** (2013).
18. Barthelmy, S. D. *et al.* GRB 130603B: Swift-BAT refined analysis. *GCN Circ.* **14741** (2013).
19. Golenetskii, S. *et al.* Konus-Wind observation of GRB 130603B. *GCN Circ.* **14771** (2013).
20. Levan, A. J. *et al.* GRB 130603B: WHT optical afterglow candidate. *GCN Circ.* **14742** (2013).
21. Thoene, C. C., de Ugarte Postigo, A., Gorosabel, J., Tanvir, N. & Fynbo, J. P. U. GRB 130603B: short GRB afterglow spectrum from GTC. *GCN Circ.* **14744** (2013).
22. Cucchiara, A. *et al.* Gemini spectroscopy of the short GRB 130603B afterglow and host. *Astrophys. J.* (submitted); preprint at <http://arxiv.org/abs/1306.2028> (2013).
23. Hjorth, J. *et al.* GRB 050509B: constraints on short gamma-ray burst models. *Astrophys. J.* **630**, L117–L120 (2005).
24. Berger, E., Fong, W. & Chornock, R. Smoking gun or smoldering embers? A possible r-process kilonova associated with the short-hard GRB 130603B. *Astrophys. J.* (submitted); <http://arxiv.org/abs/1306.3960> (2013).
25. Granot, J. & Sari, R. The shape of spectral breaks in gamma-ray burst afterglows. *Astrophys. J.* **568**, 820–829 (2002).
26. Aasi, J. *et al.* Prospects for localization of gravitational wave transients by the Advanced LIGO and Advanced Virgo Observatories. *Living Rev. Relativ.* (submitted); preprint at <http://arxiv.org/abs/1304.0670> (2013).
27. Rowlinson, A. *et al.* Discovery of the afterglow and host galaxy of the low-redshift short GRB 080905A. *Mon. Not. R. Astron. Soc.* **408**, 383–391 (2010).
28. Fong, W. *et al.* A jet break in the X-ray light curve of short GRB 111020A: implications for energetics and rates. *Astrophys. J.* **756**, 189–200 (2012).
29. Evans, P. A. *et al.* Methods and results of an automatic analysis of a complete sample of Swift-XRT observations of GRBs. *Mon. Not. R. Astron. Soc.* **397**, 1177–1201 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was based partly on observations obtained at the Gemini Observatory, which is operated by the Association of Universities for Research in Astronomy, Inc., under a cooperative agreement with the US National Science Foundation on behalf of the Gemini partnership: the National Science Foundation (United States), the National Research Council (Canada), CONICYT (Chile), the Australian Research Council (Australia), the Ministério da Ciência e Tecnologia (Brazil) and SECYT (Argentina). This work was also partly based on observations made using ESO telescopes at the Paranal Observatory. The UKIRT is operated by the Joint Astronomy Centre on behalf of the UK Science and Technology Facilities Council. The UKIRT/WFCAM data used here were pipeline-processed by the Cambridge Astronomical Survey Unit. The HST data were obtained under programme GO/DD 13497. We thank the Space Telescope Science Institute director for approving Director's Discretionary Time observations, and staff, particularly D. Taylor, for expediting them. The Dark Cosmology Centre is funded by the Danish National Research Foundation. This work made use of data supplied by the UK Swift Science Data Centre at the University of Leicester. We acknowledge discussions with A. de Ugarte Postigo and D. Watson.

Author Contributions N.R.T. led the writing of the HST proposal, performed the final photometric analysis of the HST data and took primary responsibility for writing the text of the paper. A.J.L. contributed to all aspects of the observations and planning, particularly the collation of photometry and the creation of Fig. 2. A.S.F. and R.A.H. took primary responsibility for the detailed planning of the observations and the processing of the HST imaging. J.H., K.W. and R.L.T. contributed to planning the observing and analysis strategies. All authors contributed to the writing of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.R.T. (nrt3@le.ac.uk).

A rigid and weathered ice shell on Titan

D. Hemingway¹, F. Nimmo¹, H. Zebker² & L. Iess³

Several lines of evidence suggest that Saturn's largest moon, Titan, has a global subsurface ocean beneath an outer ice shell 50 to 200 kilometres thick^{1–4}. If convection^{5,6} is occurring, the rigid portion of the shell is expected to be thin; similarly, a weak, isostatically compensated shell has been proposed^{7,8} to explain the observed topography. Here we report a strong inverse correlation between gravity³ and topography⁹ at long wavelengths that are not dominated by tides and rotation. We argue that negative gravity anomalies (mass deficits) produced by crustal thickening at the base of the ice shell overwhelm positive gravity anomalies (mass excesses) produced by the small surface topography, giving rise to this inverse correlation. We show that this situation requires a substantially rigid ice shell with an elastic thickness exceeding 40 kilometres, and hundreds of metres of surface erosion and deposition, consistent with recent estimates from local features^{10,11}. Our results are therefore not compatible with a geologically active, low-rigidity ice shell. After extrapolating to wavelengths that are controlled by tides and rotation, we suggest that Titan's moment of inertia may be even higher (that is, Titan may be even less centrally condensed) than is currently thought¹².

Three solutions for Titan's low-order gravity field^{13,12} have been obtained via radio tracking of the Cassini spacecraft, with good constraints up to spherical harmonic degree 3. Topography data have been obtained through a combination of radar altimetry and analysis of the overlapping portions of Cassini radar images¹³. Spherical harmonic models^{9,14} have been generated to represent the topography globally. Because of data gaps, the coefficients derived can depend on the maximum degree allowed for the harmonic expansion, but are stable up to degree 6 (Supplementary Information).

The degree-3 signal is useful because, although it is small compared to those at degrees 2 and 4, it is not directly affected by tides, rotation or tidal heating⁷. Figure 1 illustrates the negative correlation between the degree-3 gravity (Fig. 1a) and topography (Fig. 1b) signals for one set of gravity and topography data. We employ an admittance analysis^{15,16}, which measures the gravity-to-topography ratio in a way that is not biased by uncorrelated signals in the gravity data (Supplementary Information). For example, there may be gravity anomalies sourced from the deeper interior, but they should not show a strong correlation with surface topography. The degree-3 admittance appears to be substantially negative, and a Monte Carlo analysis shows that this result is robust to the estimated uncertainties (Fig. 2).

Negative admittances are rare. They can occur in convecting systems with strong viscosity stratification¹⁷, but it is unclear why Titan's ice shell would have such layering, and the shell is sufficiently thin that—as for Enceladus¹⁸—convective features should have much shorter horizontal length scales than degree 3.

Titan's long-wavelength surface topography probably arises from small (~1%) variations in shell thickness⁷, which may be related to tidal heating (Supplementary Information). Such variations in shell thickness can give rise to negative admittances only if the shell is rigid and loading occurs primarily from below, leading to roots at the base of the ice shell that are large compared with the isostatic case (Supplementary Information). Here we model this situation, taking

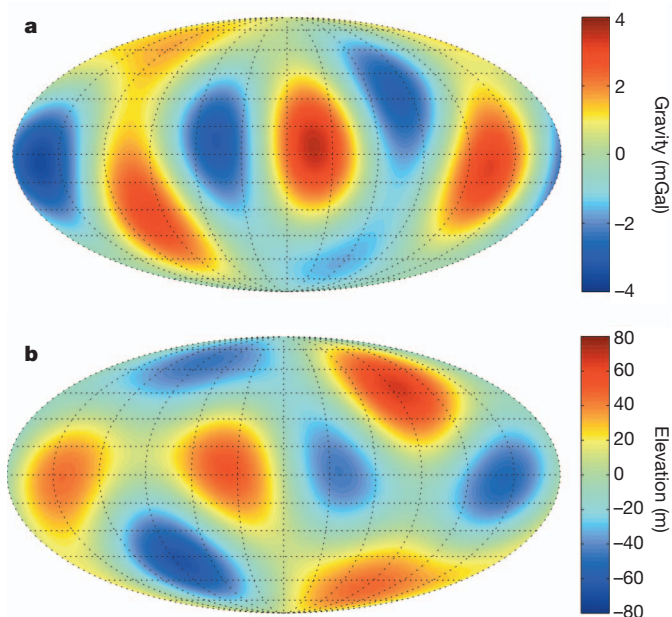


Figure 1 | Titan's degree-3 gravity and topography. **a**, Gravity field derived from the potential coefficients of the SOL1a gravity field representation of ref. 3 (multi-arc analysis, 3×3 gravity field). **b**, Spherically referenced topography based on degree-6 harmonic expansion⁹ (Supplementary Information). The two signals display a strong negative correlation (with correlation coefficient $\gamma = -0.61$) and give rise to an admittance of -32 mGal km^{-1} , based on a Monte Carlo analysis accounting for the uncertainty in the two signals (Supplementary Information). Maps are shown in Mollweide projection, centred on the anti-Saturnian point (180° longitude).

into account loading from both the top (for example, sedimentation or erosion at the surface) and bottom (for example, freezing or melting at the base of the ice shell), and calculate the resulting deformation using the equations of flexure on a thin elastic shell^{19–21}.

If the observed surface topography h and the amount of top loading d_t are specified, and if top and bottom loads are in phase, it can be shown (Supplementary Information) that the degree- l admittance is given by

$$Z(l) = \frac{l+1}{2l+1} 4\pi G \rho_{\text{crust}} \left[1 - \left(\frac{1 - \frac{d_t}{h}}{C} + \frac{d_t}{h} \right) \left(1 - \frac{d}{R} \right)^l \right] \quad (1)$$

where d is the mean ice-shell thickness, R is Titan's mean radius and ρ_{crust} is the crustal (that is, ice shell) density. The quantity d_t is the amount of material added or subtracted at the top of the shell: negative values of d_t indicate removal of material (erosion) from pre-existing highs and deposition at lows. Here C is the compensation factor at degree l ; when $C = 1$, the topography is fully compensated (isostatic) and admittance is necessarily positive. When $C < 1$, the load is partly

¹Department of Earth and Planetary Sciences, University of California Santa Cruz, 1156 High Street, Santa Cruz, California 95064, USA. ²Departments of Geophysics and Electrical Engineering, Stanford University, Stanford, California 94305, USA. ³Dipartimento di Ingegneria Meccanica e Aerospaziale, Università La Sapienza, 00184 Rome, Italy.

Topography	Degree-3 admittance (mGal km ⁻¹)	Gravity
	-40 -20 0 +20	
Degree ≤ 4		SOL1a
		SOL1b
		SOL2
Degree ≤ 5		SOL1a
		SOL1b
		SOL2
Degree ≤ 6		SOL1a
		SOL1b
		SOL2

Figure 2 | Admittance estimates for nine sets of gravity and topography data. Three distinct gravity models³ and three distinct topography models⁹ (truncated at harmonic degrees 4, 5 and 6) were used to produce a total of nine separate admittance estimates. Each estimate is based on a Monte Carlo analysis in which the admittance was computed for each of 100,000 distinct sets of gravity and topography coefficients, distributed according to the 1σ uncertainties in the model coefficients. For each Monte Carlo analysis, diamonds are plotted at the mean admittance estimate, and 1σ error bars are shown to represent the distribution. To the right of the error bars are the mean \pm standard deviation of the admittance estimates in milligals per kilometre, followed by the correlation coefficients γ in parentheses.

supported by elastic flexure of the ice shell and admittance may be negative for sufficiently small values of C or sufficiently negative values of d_t/h . In the case of degree 3, and assuming reasonable ice-shell properties (Supplementary Information), C is so large that admittance cannot be negative unless d_t/h is negative (that is, when erosion has occurred in areas of positive topography).

The black line plotted in Fig. 3a indicates the admittance corresponding to Fig. 1 and shows that more than 100 m of surface erosion is required even for very large (~ 400 km) elastic thicknesses. If the elastic thickness is less than 40 km, more than ~ 1 km of erosion is required. Figure 3b plots combinations of d , T and d_t that satisfy the observed admittance. For an ice-shell thickness of 200 km, for example, more than 200 m of erosion are required. As the elastic layer decreases in thickness, there is an increase in both the implied amount of erosion, and the absolute uncertainty in that estimate (Supplementary Information). Using $Z(3) = -32 \pm 11$ mGal km⁻¹, the uncertainty in the admittance estimate produces approximately $\pm 30\%$ uncertainty in the erosion estimate (where one galileo is defined as one centimetre per second squared).

The survival of large-impact basin rims and other topographic features^{10,11} rules out erosion amplitudes greater than ~ 1 km, implying a shell elastic thickness greater than 40 km (Fig. 3b). This high rigidity could be the result of a cold (ammonia-rich) subsurface ocean^{22,23}, a clathrate-rich shell⁴ or a low heat flux from the interior (Supplementary Information). The addition of a rigid shell would slightly reduce the k_2 tidal Love number, but not enough to conflict with the measured value³ (Supplementary Information). The implied elastic thickness rules out a vigorously convecting shell, limits the potential for widespread cryovolcanism^{22,24}, and permits the survival of lateral shell thickness variations (Supplementary Information). To generate the observed topography, our model requires shell thickness variations at degree 3; possible sources include tidal heating in a laterally heterogeneous shell²⁵, and redistribution of material via non-Newtonian lateral flow²⁶ (Supplementary Information). A rigid conductive shell is also only weakly dissipative, potentially helping to explain Titan's high present-day orbital eccentricity^{4,7}.

Our lower bound on the extent of erosion and deposition (200 m) over global length scales is compatible with estimated local amounts based on impact crater degradation^{10,11}, but somewhat larger than estimates from fluvial incision²⁷. The implied vertical erosion/deposition rate is of the order of a metre per million years, comparable to the lower end of aeolian deposition rates measured on Earth²⁸; erosion may be

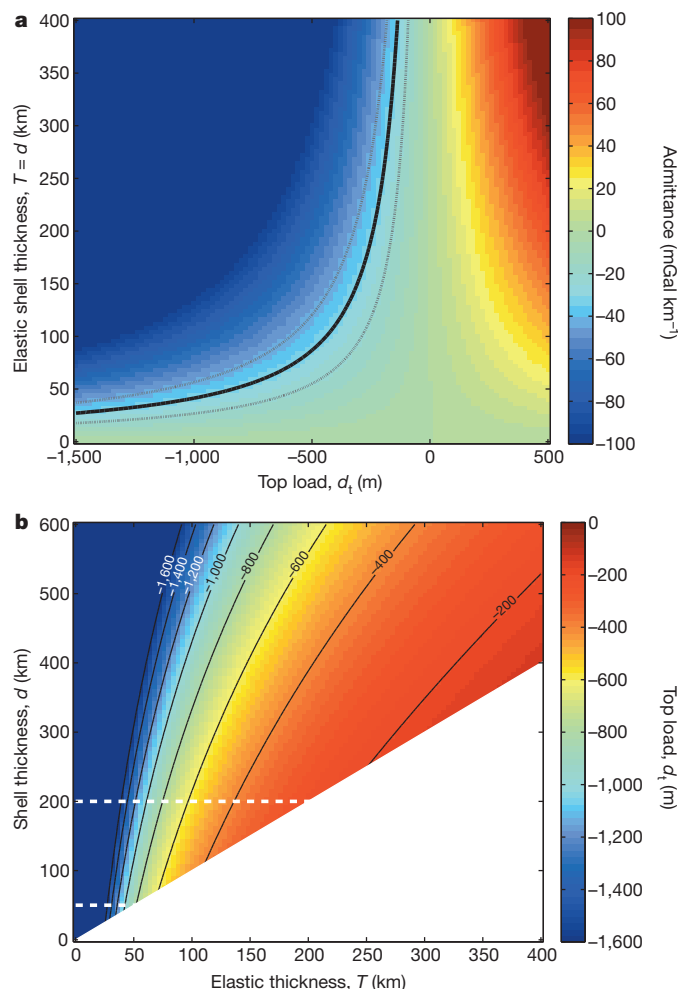


Figure 3 | Model predictions of admittance and erosion. **a**, Degree-3 admittance predicted by our model for various combinations of elastic thickness T and top load d_t , here assuming that T equals the shell thickness d . Larger values of T correspond to smaller values of C (that is, less compensation). Negative top load indicates erosion at topographic highs and deposition at topographic lows. The solid black line indicates the admittance corresponding to Fig. 1 (-32 mGal km⁻¹); the dashed lines indicate the 1σ uncertainty on that estimate. Whereas the rheology of ice implies that $T \approx 0.5d$ (Supplementary Information), assuming that $T = d$ leads to conservative estimates of erosion. **b**, Top load amplitude, d_t , required to produce the observed degree-3 admittance for various combinations of total shell thickness d and elastic thickness T . Dashed white lines highlight the likely range of Titan's ice shell thickness, 50–200 km. Both **a** and **b** were generated assuming a degree-3 topographic amplitude h of 66 m.

occurring by physical comminution, dissolution or sublimation. Because Titan's topography is high at the equator, we predict maximum erosion occurring at the equator and sediment transport, via fluvial or aeolian processes, predominantly towards the poles. An alternative possibility is transport in the vapour phase, if the mobile material has a sublimation temperature close to that of the surface temperature.

The degree-2 gravity signal is dominated by tides and rotation, and has been used to determine Titan's moment of inertia¹². If the processes identified here are also operating at degree 2, then the measured degree-2 gravity signal is an underestimate of that caused by tides and rotation (Supplementary Information). As a result, Titan's fluid Love number may be larger than the value estimated directly from the observed gravity field, indicating that Titan may be even less centrally condensed than previously thought (Supplementary Information). This result reinforces the need to understand how so large a body could have accreted without undergoing more complete differentiation^{12,29}.

If the ice shell is sufficiently rigid (that is, the elastic layer exceeds half of the total shell thickness), then the admittance at degree 4 should also be negative. Assuming a fully elastic shell (that is, $T = d$) of thickness 200 km, we predict the degree-4 gravity field amplitude and the degree-4 admittance to be 1.4 mGal and -5 mGal km^{-1} , respectively (Supplementary Information). Future Cassini gravity fly-bys will improve the determination of the degree-4 gravity field by a factor of two and will therefore provide a partial test of this prediction. If a non-negative degree-4 admittance is observed, that may be an indication that a smaller portion (less than half) of the ice shell is elastic.

Received 8 May; accepted 20 June 2013.

- Béghin, C., Sotin, C. & Hamelin, M. Titan's native ocean revealed beneath some 45 km of ice by a Schumann-like resonance. *C. R. Geosci.* **342**, 425–433 (2010).
- Bills, B. G. & Nimmo, F. Rotational dynamics and internal structure of Titan. *Icarus* **214**, 351–355 (2011).
- Iess, L. *et al.* The tides of Titan. *Science* **337**, 457–459 (2012).
- Tobie, G., Lunine, J. I. & Sotin, C. Episodic outgassing as the origin of atmospheric methane on Titan. *Nature* **440**, 61–64 (2006).
- Mitri, G. & Showman, A. P. Thermal convection in ice-I shells of Titan and Enceladus. *Icarus* **193**, 387–396 (2008).
- Tobie, G., Grasset, O., Lunine, J. I., Mocquet, A. & Sotin, C. Titan's internal structure inferred from a coupled thermal-orbital model. *Icarus* **175**, 496–502 (2005).
- Nimmo, F. & Bills, B. G. Shell thickness variations and the long-wavelength topography of Titan. *Icarus* **208**, 896–904 (2010).
- Choukroun, M. & Sotin, C. Is Titan's shape caused by its meteorology and carbon cycle? *Geophys. Res. Lett.* **39**, 1–5 (2012).
- Zebker, H. A. *et al.* Titan's figure fatter, flatter than its gravity field. *AGU Fall Meet. abstr.* P23F–01 (2012).
- Neish, C. D. *et al.* Crater topography on Titan: implications for landscape evolution. *Icarus* **223**, 82–90 (2013).
- Moore, J. M., Howard, A. D. & Schenk, P. M. Bedrock denudation on Titan: estimates of vertical extent and lateral debris dispersion. *Lunar Planet. Sci. Conf. XXXVIII*, abstr. 1763 (2013).
- Iess, L. *et al.* Gravity field, shape, and moment of inertia of Titan. *Science* **327**, 1367–1369 (2010).
- Stiles, B. W. *et al.* Determining Titan surface topography from Cassini SAR data. *Icarus* **202**, 584–598 (2009).
- Zebker, H. *et al.* Size and shape of Saturn's moon Titan. *Science* **324**, 921–923 (2009).
- McKenzie, D. The relationship between topography and gravity on Earth and Venus. *Icarus* **112**, 55–88 (1994).
- Wieczorek, M. A. Gravity and topography of the terrestrial planets. *Treat. Geophys.* **10**, 165–206 (2007).
- Richards, M. A. & Hager, B. H. Geoid anomalies in a dynamic Earth. *J. Geophys. Res.* **89**, 5987–6002 (1984).
- Roberts, J. H. & Nimmo, F. Tidal heating and the long-term stability of a subsurface ocean on Enceladus. *Icarus* **194**, 675–689 (2008).
- Kraus, H. *Thin Elastic Shells* (Wiley, 1967).
- Turcotte, D. L., Willemann, R. J., Haxby, W. F. & Norberry, J. Role of membrane stresses in the support of planetary topography. *J. Geophys. Res.* **86**, 3951–3959 (1981).
- McGovern, P. J. *et al.* Localized gravity/topography admittance and correlation spectra on Mars: implications for regional and global evolution. *J. Geophys. Res.* **107**, 5136 (2002).
- Moore, J. M. & Pappalardo, R. T. Titan: an exogenic world? *Icarus* **212**, 790–806 (2011).
- Grasset, O., Sotin, C. & Deschamps, F. On the internal structure and dynamics of Titan. *Planet. Space Sci.* **48**, 617–636 (2000).
- Lopes, R. M. C. *et al.* Cryovolcanic features on Titan's surface as revealed by the Cassini Titan Radar Mapper. *Icarus* **186**, 395–412 (2007).
- Běhounková, M., Tobie, G., Choblet, G. & Čadež, O. Tidally-induced melting events as the origin of south-pole activity on Enceladus. *Icarus* **219**, 655–664 (2012).
- Nimmo, F. Non-Newtonian topographic relaxation on Europa. *Icarus* **168**, 205–208 (2004).
- Black, B. A., Perron, J. T., Burr, D. M. & Drummond, S. A. Estimating erosional exhumation on Titan from drainage network morphology. *J. Geophys. Res.* **117**, E08006 (2012).
- Patterson, D. B., Farley, K. A. & Norman, M. D. He-4 as a tracer of continental dust: a 1.9 million year record of aeolian flux to the west equatorial Pacific Ocean. *Geochim. Cosmochim. Acta* **63**, 615–625 (1999).
- O'Rourke, J. G. & Stevenson, D. J. Stability of ice/rock mixtures with applications to Titan. *Lunar Planet. Sci. Conf. XXXII*, abstr. 1629 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the Cassini radar science team, M. Manga, D. Stevenson, R. Pappalardo and W. McKinnon for their suggestions. Portions of this work were supported by NASA grants NNX13AG02G and NNX11AK44G.

Author Contributions F.N. initiated the effort. D.H. and F.N. developed the loading models and analysed the results. L.I. led the development of the gravity field models. H.Z. synthesized the topography models. All authors discussed the results and implications and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.H. (djheming@ucsc.edu).

Crystallites of magnetic charges in artificial spin ice

Sheng Zhang^{1*}, Ian Gilbert^{2*}, Cristiano Nisoli³, Gia-Wei Chern³, Michael J. Erickson⁴, Liam O'Brien^{4,5}, Chris Leighton⁴, Paul E. Lammert¹, Vincent H. Crespi¹ & Peter Schiffer²

Artificial spin ice¹ is a class of lithographically created arrays of interacting ferromagnetic nanometre-scale islands. It was introduced to investigate many-body phenomena related to frustration and disorder in a material that could be tailored to precise specifications and imaged directly. Because of the large magnetic energy scales of these nanoscale islands, it has so far been impossible to thermally anneal artificial spin ice into desired thermodynamic ensembles; nearly all studies of artificial spin ice have either treated it as a granular material activated by alternating fields² or focused on the as-grown state of the arrays³. This limitation has prevented experimental investigation of novel phases that can emerge from the nominal ground states of frustrated lattices. For example, artificial kagome spin ice, in which the islands are arranged on the edges of a hexagonal net, is predicted to support states with monopolar charge order at entropies below that of the previously observed pseudo-ice manifold⁴. Here we demonstrate a method for thermalizing artificial spin ices with square and kagome lattices by heating above the Curie temperature of the constituent material. In this manner, artificial square spin ice achieves unprecedented thermal ordering of the moments. In artificial kagome spin ice, we observe incipient crystallization of the magnetic charges embedded in pseudo-ice, with crystallites of magnetic charges whose size can be controlled by tuning the lattice constant. We find excellent agreement between experimental data and Monte Carlo simulations of emergent charge–charge interactions.

In the past few years, the low-temperature physics of the pyrochlore spin-ice materials has been described in terms of magnetic monopole excitations^{5–8}. In such systems, the three-dimensional pyrochlore lattice is composed of corner-sharing tetrahedra, with rare-earth-ion magnetic moments residing on the corners of the tetrahedra. The ground state of each tetrahedron consists of two spins pointing in and two spins pointing out of the tetrahedron, satisfying Pauling's 'ice rule'. If each magnetic dipole is represented as a dimer of two opposite magnetic charges, then this ice rule corresponds to charge minimization within each tetrahedron, and excitations have an effective monopolar charge.

This dimer picture can be applied generally to a range of 'ice-like' magnetic systems. For lattices such as pyrochlore with even numbers of spins around each point of convergence, local energy minimization yields perfect charge cancellation, and excitations correspondingly lead to effective monopolar charges as topological defects in an emergent gauge field. In lattices with odd spin coordination, however, the cancellation is imperfect and one obtains a lattice of effective magnetic charges even in the absence of excitation (a multipole expansion shows that Coulomb's law can approximate the interactions among these magnetic charges^{9,10}). Although all of these charges can broadly be defined as monopoles, a more restrictive definition of 'monopole' reserves that term for mobile topological defects that violate an emergent gauge field⁵. Under either definition, however, the collective dynamics of the net magnetic charges in these systems are a fascinating venue for the study

of emergent behaviour, and we avoid the term 'monopole' in the discussion below to minimize the semantic ambiguity.

Artificial spin-ice systems¹ are composed of arrays of single-domain ferromagnetic nanoislands whose magnetic moments behave like giant Ising spins and whose geometry results in competing interactions that mimic the 'ice models' of statistical mechanics¹¹. The magnetic moments of the individual nanoislands can be imaged by magnetic force microscopy (MFM), permitting the direct observation of magnetic charges^{3,12–14}. The square and kagome lattices, as shown in Fig. 1, have been the focus of most attention.

Artificial square spin ice, the first such system to be studied, consists of islands having the same coordination number as the atomic spins in naturally occurring spin-ice materials but has no residual entropy. Its two-fold degenerate ground state has proved difficult to reach^{15,16}, although some groups have achieved sizeable domains in as-grown samples³. We note that, in this system with even coordination number, magnetic charges exist only as excitations. Artificial kagome spin ice corresponds to spins arranged on a kagome lattice^{4,17}. This system has been predicted to possess a complex phase diagram, including a charge-ordered phase due to the interactions of the magnetic charges intrinsic to the odd coordination at each vertex^{9,10}. These phases have

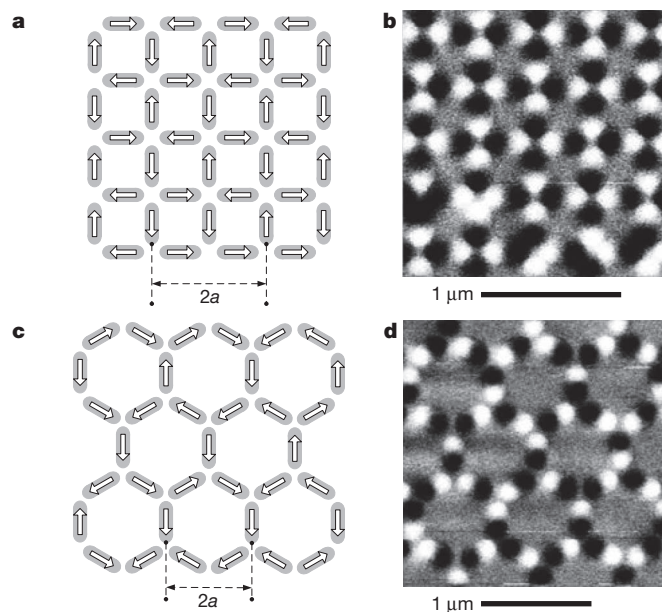


Figure 1 | The artificial spin-ice lattices used in this study. Left panels (a, c) illustrate the nanomagnet configurations used to create artificial square (a) and kagome (c) spin ice. Right panels (b, d) show MFM images of artificial square (b) and kagome (d) spin ice. The black and white spots correspond to the magnetic poles of the islands. The arrows in a and c correspond to the magnetic moments revealed by the corresponding MFM images. *a*, Lattice constant.

¹Department of Physics and Materials Research Institute, Pennsylvania State University, University Park, Pennsylvania 16802, USA. ²Department of Physics and Frederick Seitz Materials Research Laboratory, University of Illinois, Urbana, Illinois 61801, USA. ³Theoretical Division, MS B258, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. ⁴Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, Minnesota 55455, USA. ⁵Thin Film Magnetism Group, Department of Physics, Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK.

*These authors contributed equally to this work.

only begun to be explored experimentally¹⁸, and have not been visualized directly as experimental evidence for the collective behaviour of magnetic charges and their usefulness as degrees of freedom in low-energy collective physics. We use here a thermal annealing technique to achieve the first experimental validation of the predicted charge-ordered phase: incipient magnetic charge ordering in artificial kagome spin ice.

The current gap between theory and experiment is in part a consequence of the present limits on our ability to place the magnetic moments of artificial spin ice in the desired collective thermodynamic state. Although a.c. demagnetization can produce well defined and predictable out-of-equilibrium statistical ensembles of different entropy, it cannot approach the two-fold degenerate ground state of square ice¹⁵. In the case of artificial kagome spin ice, a.c. demagnetization could easily access the kagome ice I phase (in which the islands obey the two-in/one-out or one-in/two-out pseudo-ice rule), probably because of that state's massive entropy. Yet a.c. demagnetization failed to produce any of the predicted ordered phases governed by long-range interactions.

Recently a few alternative ways to thermalize artificial spin ice have been developed. Morgan and co-workers examined^{3,19} the as-grown state of artificial square spin ice and discovered that thermalization during the first stages of growth is preserved, leaving the sample in a very low energy state. This single-shot approach represents the best energy minimization available to date for artificial spin ice, although its reliance on the initial conditions of growth limits the degree of control which can be obtained by this method. Kapaklis *et al.* fabricated²⁰ artificial spin ice in a material that has a Curie temperature (T_C) of 230 K, below room temperature. Magneto-optical Kerr effect (MOKE) measurements of remanent states revealed evidence for the creation of thermal excitations slightly below the Curie temperature. Although this work in some sense represents the successful thermalization of artificial spin ice, it focused on the remanent state (rather than the true ground state), and the MOKE measurements were sensitive only to the global properties of the samples, so the exact microstates of the samples were not determined. Several groups^{21–23}, inspired by these experiments, have carried out simulations of thermalized artificial spin-ice arrays, but post-growth thermalization of artificial spin ice into its zero-field ground state has not yet been experimentally demonstrated.

The results of Morgan and co-workers demonstrated that thermalization of artificial spin ice is possible, inspiring the present work in which we achieve such thermalization on demand and then use it to explore the long-range collective physics. We fabricated square and kagome arrays of permalloy ($\text{Ni}_{80}\text{Fe}_{20}$) nanoislands on a silicon nitride substrate using an electron beam lithography lift-off process described elsewhere¹. The islands are rectangular with semicircular ends ('stadium-shaped'), nominally 220 nm long by 80 nm wide and 25 nm high. The single-domain character of the islands was confirmed by MFM (right panels of Fig. 1). We studied arrays with a wide range of lattice constants (320–880 nm for square, and 260–1,050 nm for kagome) so as to have a range of magnetostatic interaction strengths. The arrays were then polarized in a strong magnetic field at room temperature, heated in vacuum in nominally zero magnetic field to an annealing temperature above the islands' Curie temperature (540–545 °C as measured by vibrating sample magnetometry), and then allowed to cool slowly to room temperature. Square and kagome arrays of all lattice constants were fabricated together on chips so that all the arrays on a chip shared the same annealing protocol. Results are shown for an annealing temperature of 545 °C, but qualitatively consistent results were obtained on multiple samples annealed just above the Curie temperature. Importantly, by judicious choice of substrate materials and careful scanning of the relevant temperature range in annealing experiments, we were able to identify a temperature window in which effective annealing can be performed while avoiding film/substrate interdiffusion and lateral degradation of the islands. Details of sample fabrication and this annealing process can be found in Methods, Supplementary Information and

Supplementary Figs 1–3. Our expectation was that the island moments would align into a minimum-energy state on cooling through the Curie temperature, and that this state would be determined by the lattice geometry, that is, the artificial spin ice would be thermalized.

After annealing, we collected MFM images of each array, which allowed us to examine the thermalized state of the moment arrangements. Images of the annealed artificial square spin-ice arrays revealed extensive domains of ground-state ordering (Fig. 2). The ground state consists of an antiferromagnetic ordering on the vertical and horizontal sublattices. The large size of the domains (in many cases larger than the 10–20- μm -wide MFM images) and the paucity of isolated monopolar charge excitations, especially in the samples with the smallest lattice constants, underscore the effectiveness of our annealing process at minimizing the magnetostatic energy of the arrays. In the arrays with smallest lattice constant, only the handful of islands located on the isolated domain walls are not in the ground state (Supplementary Fig. 4). Approximating the islands' magnetization as proportional to the square root of $T_C - T$ (where T is the physical sample temperature), we found the excitations of the 400-nm square lattice fit well to a Boltzmann distribution corresponding to a temperature of $T \approx 540$ °C, as shown in Supplementary Fig. 6.

To gain further insight into the development of these large ground-state domains in annealed artificial square spin ice, we extracted the pairwise correlations between islands from our MFM images. The correlation between two islands is +1 (–1) if the islands' magnetic moments are aligned to minimize (maximize) the dipolar interaction energy¹. A plot of correlation as a function of island separation is shown in Fig. 3. The distance at which the correlation falls to $1/e \approx 0.368$ corresponds to a rough estimate of the domain size²⁴. Because the dipolar interaction between islands is largest for the arrays with the smallest lattice constants, the correlations in these arrays decay the most slowly. This can be qualitatively confirmed by examining the MFM images of Fig. 2: smaller lattice constants produce larger ground-state domains. The correlation between islands at the largest lattice spacings is insignificant beyond one or two lattice constants.

The ground-state ordering observed in our artificial square spin ice highlights a significant difference between artificial square spin ice and

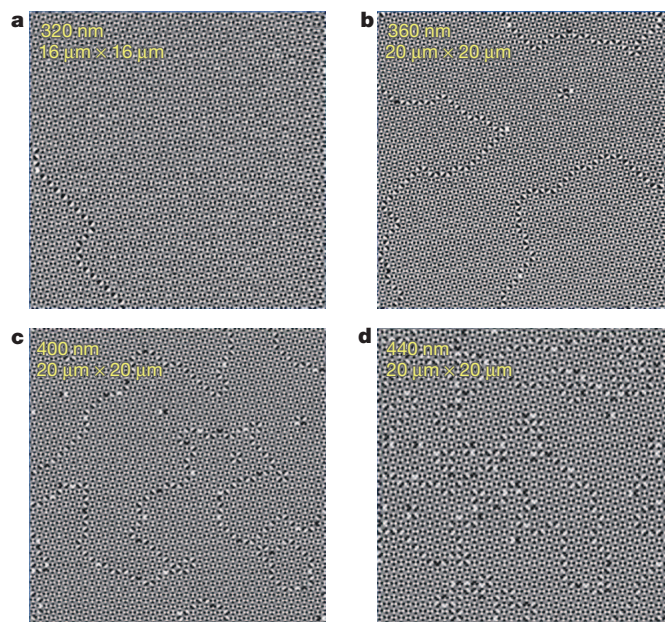


Figure 2 | MFM images of annealed artificial square spin ice. a–d, The lattice constants are given at top left, and are 320 nm (a), 360 nm (b), 400 nm (c) and 440 nm (d). The image size is also indicated at top left. Notice the decrease in domain size with increasing lattice constant, a trend also revealed in Fig. 3 by correlations between islands. The ordering of moments is in stark contrast with the disordered state resulting from a.c. demagnetization^{1,2}.

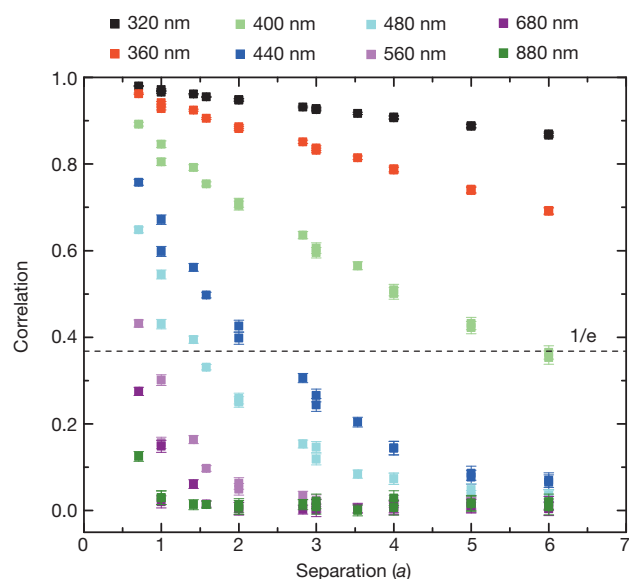


Figure 3 | Magnitude of the correlation between islands as a function of island separation for annealed artificial square spin ice of all lattice spacings. The different curves correspond to arrays of different lattice constants, as indicated in the key at the top. The intersection of each curve with the horizontal dashed line at $1/e$ can be taken as an estimate of the radius of an average domain. Error bars (± 1 s.d.) are calculated assuming the island pairs have a binomial distribution (because they can either favourably or unfavourably aligned).

natural spin-ice materials: this two-dimensional lattice of artificial spin ice is less frustrated than the three-dimensional lattice of spin-ice materials, so all the nearest-neighbour (islands meeting at 90° to each other at a vertex) interactions can be satisfied. The ability of thermally annealed artificial square spin ice to access its ground state experimentally allows a more direct connection with simulations^{21–23} than in natural spin-ice materials, where no ordering is observed experimentally down to the lowest temperatures.

In contrast to artificial square spin ice, artificial kagome spin ice does not possess a ground state that satisfies all nearest-neighbour interactions—each vertex of three islands will have a net magnetic charge of ± 1 when obeying the pseudo-ice rule. Nonetheless, the long-range tail of the dipolar interaction can further reduce the entropy, and multipole expansion^{5,9} shows that the long-range interactions can be subsumed into a magnetic charge picture. Theoretical work^{9,10} predicts a transition from the kagome ice I phase, which consists of a random arrangement of these charges in a pseudo-ice manifold, to a charge-ordered state (kagome ice II) on a bipartite hexagonal lattice. Weak signatures of the long-range nature of the dipolar interaction were reported previously in terms of a slight suppression or enhancement of certain spin correlations^{4,25} or a small reduction of the overall entropy²⁶. Experimental observation of interacting magnetic charges collectively organized into the kagome ice II state has, however, remained elusive.

In Fig. 4a, we show an MFM image of an annealed kagome ice sample. For all such annealed samples with small lattice constant (260 nm and 300 nm), all of the vertices obey the kagome ice rule, so the system is at least in the kagome ice I state (Supplementary Fig. 5). Furthermore, for the 260-nm lattice constant shown in Fig. 4a, we also see an incipient charge ordering, with formation of crystallites of ordered charges (note that the ordering does not require the ordering of the individual island dipoles; even a perfectly crystallized state has a residual entropy^{9,10}). These domains are apparent in Fig. 4b, where we colour charge-ordered domains red and blue (the coloured dots correspond to vertices belonging to the two degenerate ordering configurations).

We extracted the entropy of the artificial kagome spin-ice arrays (Fig. 5c inset) using a method based on conditional probability²⁶. We

measure the entropy (S) as bits per spin, so that a random configuration has $S = 1$, the kagome ice I state has $S \approx 0.724$, and the kagome ice II state is predicted to have $S \approx 0.15$ (ref. 26). Although the thermally annealed artificial kagome spin-ice array with the smallest lattice constant has an entropy substantially above 0.15, its entropy falls well below that expected for kagome ice I, unlike artificial kagome spin ice treated with an a.c. demagnetization protocol. This supports the picture of charges beginning to order into the kagome ice II state.

Although the exact form of magnetic interactions between the islands depends on several geometric parameters of the artificial spin-ice arrays, the nature of the charge ordering transition is independent of these microscopic details. The twofold degeneracy of the charge-ordered states indicates a phase transition belonging to the Ising universality class. This observation allows us to compare the experimental results with simulations based on a kagome ice model with short-range nearest-neighbour monopolar charge interactions. To assess the viability of a description of low-entropy states in terms of interacting magnetic charges, we have performed Monte Carlo simulations (details in Supplementary Information and Supplementary Figs 7–10) and extracted charge–charge correlation functions (Fig. 5a and b). Overall agreement is demonstrated by plotting the extracted entropy as a function of nearest-neighbour charge–charge correlation, C_1 (Fig. 5c). As a whole, these results confirm that the annealing protocol produces thermal ensembles that can be effectively described in terms of an emergent picture of interacting magnetic charges. Note that the observed crystallization is emergent because of the role played by magnetic charges embedded in the pseudo-ice manifold, and this emergent nature explains the reduction in entropy.

In summary, we have demonstrated the possibility of thermally annealing artificial spin ice and thus accessing the subtle submanifolds that emerge in this system. Unlike a.c. demagnetization, our protocol produces truly thermalized microscopic ensembles. In artificial square spin ice, it produces a ground state of unprecedented quality. For artificial kagome spin ice we report direct visualization of magnetic charge crystallization. We thereby validate a theoretically predicted

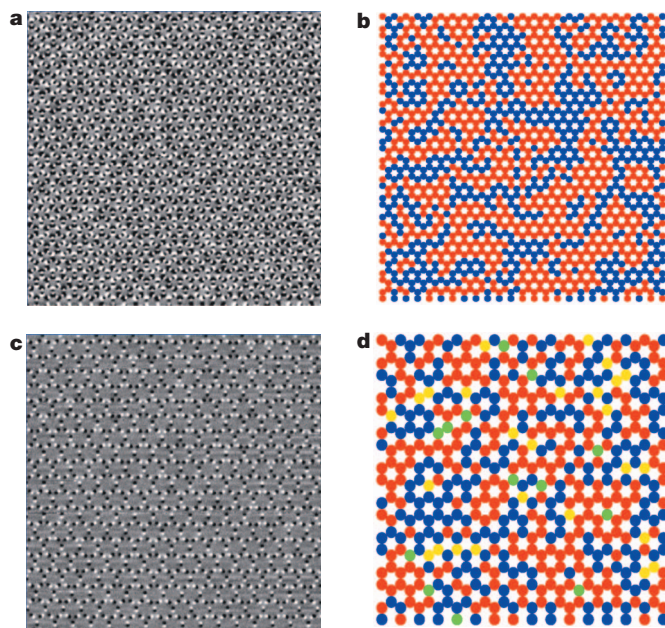


Figure 4 | Emergent domains of ordered magnetic charges in artificial kagome spin ice. a, c, MFM images of thermally annealed artificial kagome spin ice with lattice parameter of 260 nm (a) and 490 nm (c). b, d, Maps of the distribution of magnetic charge in a and c, respectively. Red and blue dots correspond to vertices belonging to each of the two degenerate magnetic charge-ordered states described in the text. Green and yellow dots represent the $+3$ and -3 excitations. Supplementary Fig. 11 reproduces these maps but also includes vectors representing the magnetic moments of the islands comprising the lattices.

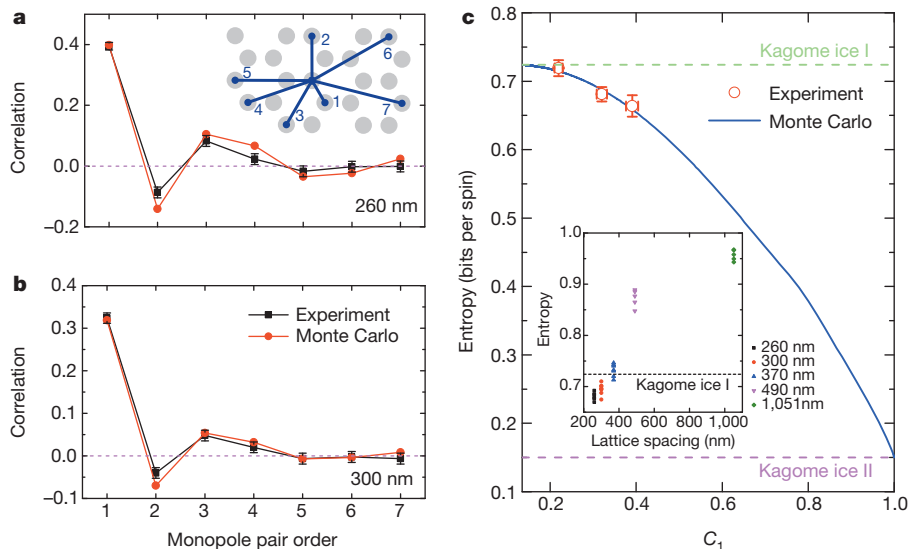


Figure 5 | Charge-charge correlation functions of annealed artificial kagome spin ice. **a, b,** Measured charge-charge correlation (black squares) for lattice constants of 260 nm (**a**) and 300 nm (**b**). Error bars are analogous to those in Fig. 3. Correlation functions obtained from corresponding Monte Carlo simulations are also shown (red circles). The inset in **a** defines the monopole pair order. **c,** Extracted and simulated entropy as a function of

emergent picture, which subsumes the complexities of the long-range physics into an intuitive picture of interacting magnetic charges, bona fide dynamical objects whose interaction explains the thermodynamics of the system as it approaches a charge-ordered state. The key point is that an effective, simplified description (interacting magnetic charges) captures the essential elements of the system in a smaller number of effective degrees of freedom than the full description (all of the island moments).

This method will now allow the true thermalization of different moment configurations—including lithographically introduced defects and novel array geometries^{27,28}. The technique will add to very recent progress such as the study of thermally induced dynamics in clusters of extremely thin islands²⁹ and the incomplete ground state ordering observed in a thermal anneal of an artificial square ice lattice composed of a material in which the island moments' Curie temperature was tuned to be lower than that of the standard permalloy that we studied³⁰. Applied magnetic fields during thermal annealing will allow direct comparison of these systems to the temperature/magnetic-field phase diagram of pyrochlore spin-ice materials. Thermal annealing also invites meaningful comparison of artificial spin ice to traditional Monte Carlo simulations, allowing artificial spin ice to once again serve as a bridge between theory and experiment for frustrated magnetic materials.

METHODS SUMMARY

Our nanomagnet arrays were fabricated using a bilayer resist stack similar to that used in previous studies¹. A silicon substrate coated with 2,000 Å low-pressure CVD silicon nitride was used to inhibit film-substrate interdiffusion, thereby mitigating any degradation in magnetic properties during annealing. The arrays were then defined with a Leica EBPG-5HR electron beam lithography writer. Permalloy ($\text{Ni}_{80}\text{Fe}_{20}$) films of thickness 25 nm were subsequently electron-beam-evaporated at room temperature at 0.5 Å s^{-1} and a base pressure of 10^{-10} torr from a 99.95% purity source. A 3-nm Al protective cap layer was deposited to prevent permalloy oxidation. Film thickness was calibrated using grazing incidence X-ray reflectivity and monitored during deposition using a quartz crystal monitor. Scanning electron microscopy, atomic force microscopy and X-ray diffraction indicate a polycrystalline film with approximately 21 nm grain size and 0.7 nm roughness over a $2 \times 2 \mu\text{m}$ area. Lift-off was done in the lift-off solvent Remover PG (MicroChem Corp.) at 65°C , followed by sonication in acetone. Samples were fully magnetized by an applied magnetic field of 2,000 Oe along the (1, 1) direction of square arrays and then annealed in a magnetically shielded vacuum chamber (base pressure

nearest-neighbour charge-charge correlation, C_1 . Inset, entropy as a function of lattice spacing. Entropy error bars are ± 1 s.d., extracted from several MFM images, shown as separate points in the inset. At the smallest lattice spacing, the entropy falls significantly below that expected for kagome ice I, suggesting an incipient transition to the lower-entropy charge-ordered phase.

10^{-7} torr) in magnetic fields below 0.1 G. Samples were heated at a rate of $20^\circ\text{C min}^{-1}$ and held at an annealing temperature of 545°C for 15 min before being cooled at 1°C min^{-1} . MFM images revealed the net magnetic moment of the arrays was drastically reduced by the annealing process, confirming the effectiveness of the annealing process. Images of the magnetic moments of the individual islands were acquired by MFM, and we used computerized automatic image recognition routines to extract the exact microstate of the arrays from the images. Full Methods are available in Supplementary Information.

Received 21 January; accepted 14 June 2013.

- Wang, R. F. *et al.* Artificial 'spin ice' in a geometrically frustrated lattice of nanoscale ferromagnetic islands. *Nature* **439**, 303–306 (2006).
- Ke, X. *et al.* Energy minimization and AC demagnetization in a nanomagnet array. *Phys. Rev. Lett.* **101**, 037205 (2008).
- Morgan, J. P., Stein, A., Langridge, S. & Marrows, C. H. Thermal ground-state ordering and elementary excitations in artificial magnetic square ice. *Nature Phys.* **7**, 75–79 (2011).
- Qi, Y., Brintlinger, T. & Cumings, J. Direct observation of the ice rule in an artificial kagome spin ice. *Phys. Rev. B* **77**, 094418 (2008).
- Castelnovo, C., Moessner, R. & Sondhi, S. L. Magnetic monopoles in spin ice. *Nature* **451**, 42–45 (2008).
- Fennell, T. *et al.* Magnetic Coulomb phase in the spin ice $\text{Ho}_2\text{Ti}_2\text{O}_7$. *Science* **326**, 415–417 (2009).
- Morris, D. J. *et al.* Dirac strings and magnetic monopoles in the spin ice $\text{Dy}_2\text{Ti}_2\text{O}_7$. *Science* **326**, 411–414 (2009).
- Ryzhkin, I. A. Magnetic relaxation in rare-earth oxide pyrochlores. *J. Exp. Theor. Phys.* **101**, 481–486 (2005).
- Möller, G. & Moessner, R. Magnetic multipole analysis of kagome and artificial spin-ice dipolar arrays. *Phys. Rev. B* **80**, 140409(R) (2009).
- Chern, G.-W., Mellado, P. & Tchernyshyov, O. Two-stage ordering of spins in dipolar spin ice on the kagome lattice. *Phys. Rev. Lett.* **106**, 207202 (2011).
- Baxter, R. J. *Exactly Solved Models in Statistical Mechanics* (Academic, 1982).
- Pollard, S. D., Volkov, V. & Zhu, Y. Propagation of magnetic charge monopoles and Dirac flux strings in an artificial spin-ice lattice. *Phys. Rev. B* **85**, 180402(R) (2012).
- Ladak, S., Read, D. E., Perkins, G. K., Cohen, L. F. & Branford, W. R. Direct observation of magnetic monopole defects in an artificial spin-ice system. *Nature Phys.* **6**, 359–363 (2010).
- Mengotti, E. *et al.* Real-space observation of emergent magnetic monopoles and associated Dirac strings in artificial kagome spin ice. *Nature Phys.* **7**, 68–74 (2011).
- Nisoli, C. *et al.* Effective temperature in an interacting vertex system: theory and experiment on artificial spin ice. *Phys. Rev. Lett.* **105**, 047205 (2010).
- Remhof, A. *et al.* Magnetostatic interactions on a square lattice. *Phys. Rev. B* **77**, 134409 (2008).
- Möller, G. & Moessner, R. Artificial square ice and related dipolar nanoarrays. *Phys. Rev. Lett.* **96**, 237202 (2006).

18. Branford, W. R., Ladak, S., Read, D. E., Zeissler, K. & Cohen, L. F. Emerging chirality in artificial spin ice. *Science* **335**, 1597–1600 (2012).
19. Morgan, J. *et al.* Real and effective thermal equilibrium in artificial square spin ices. *Phys. Rev. B* **87**, 024405 (2013).
20. Kapaklis, V. *et al.* Melting artificial spin ice. *New J. Phys.* **14**, 035009 (2012).
21. Budrikis, Z. *et al.* Domain dynamics and fluctuations in artificial square ice at finite temperatures. *New J. Phys.* **14**, 035014 (2012).
22. Silva, R. C. *et al.* Thermodynamics of elementary excitations in artificial magnetic square ice. *New J. Phys.* **14**, 015008 (2012).
23. John Greaves, S. & Muraoka, H. Formation of thermally induced ground states in two-dimensional square spin ices. *J. Appl. Phys.* **112**, 043909 (2012).
24. Newman, M. E. J. & Barkema, G. T. *Monte Carlo Methods in Statistical Physics* (Clarendon, 1999).
25. Rougemaille, N. *et al.* Artificial kagome arrays of nanomagnets: a frozen dipolar spin ice. *Phys. Rev. Lett.* **106**, 057209 (2011).
26. Lammert, P. E. *et al.* Direct entropy determination and application to artificial spin ice. *Nature Phys.* **6**, 786–789 (2010).
27. Morrison, M. J., Nelson, T. R. & Nisoli, C. Unhappy vertices in artificial spin ice: degeneracy from vertex-frustration. *New J. Phys.* **15**, 045009 (2013).
28. Chern, G.-W., Morrison, M. J. & Nisoli, C. Engineering degeneracy: a critical ground state for artificial spin ice. Preprint at <http://arxiv.org/abs/1210.8377> (2012).
29. Farhan, A. *et al.* Exploring hyper-cubic energy landscapes in thermally active finite artificial spin-ice systems. *Nature Phys.* **9**, 375–382 (2013).
30. Porro, J. M. *et al.* Exploring thermally induced states in square artificial spin-ice arrays. *New J. Phys.* **15**, 055012 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements This project was funded by the US Department of Energy, Office of Basic Energy Sciences, Materials Sciences and Engineering Division under grant no. DE-SC0005313. Lithography was performed with the support of the National Nanotechnology Infrastructure Network. The work of C.N. and G.-W.C. was carried out under the auspices of the US Department of Energy at LANL under contract no. DE-AC52-06NA253962. Work at the University of Minnesota was supported by the NSF MRSEC under award DMR-0819885 and EU Marie Curie IOF project no. 299376. Certain theory elements were supported by the NSF MRSEC under award DMR-0820404.

Author Contributions P.S., C.L. and V.H.C. initiated this study and supervised the experiments, data analysis and simulations. Artificial spin-ice array fabrication, measurement and data analysis was performed by S.Z. and I.G. Thin film deposition and characterization was performed by L.O'B. and M.J.E. The thermal annealing process was refined and performed by L.O'B., M.J.E. and C.L. Simulations, some data analysis and theoretical interpretation were provided by C.N., G.-W.C. and P.E.L. I.G., C.N. and P.S. wrote the paper with input from all the co-authors. S.Z. and I.G. contributed equally to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.S. (pschiffe@illinois.edu).

Late Miocene threshold response of marine algae to carbon dioxide limitation

Clara T. Bolton¹ & Heather M. Stoll¹

Coccolithophores are marine algae that use carbon for calcification and photosynthesis. The long-term adaptation of these and other marine algae to decreasing carbon dioxide levels during the Cenozoic era¹ has resulted in modern algae capable of actively enhancing carbon dioxide at the site of photosynthesis. This enhancement occurs through the transport of dissolved bicarbonate (HCO_3^-) and with the help of enzymes whose expression can be modulated by variable aqueous carbon dioxide concentration, $[\text{CO}_2]$, in laboratory cultures^{2,3}. Coccolithophores preserve the geological history of this adaptation because the stable carbon and oxygen isotopic compositions of their calcite plates (coccoliths), which are preserved in the fossil record, are sensitive to active carbon uptake and transport by the cell. Here we use a model of cellular carbon fluxes and show that at low $[\text{CO}_2]$ the increased demand for HCO_3^- at the site of photosynthesis results in a diminished allocation of HCO_3^- to calcification, which is most pronounced in larger cells. This results in a large divergence between the carbon isotopic compositions of small versus large coccoliths only at low $[\text{CO}_2]$. Our evaluation of the oxygen and carbon isotope record of size-separated fossil coccoliths reveals that this isotopic divergence first arose during the late Miocene to the earliest Pliocene epoch (about 7–5 million years ago). We interpret this to be a threshold response of the cells' carbon acquisition strategies to decreasing $[\text{CO}_2]$. The documented coccolithophore response is synchronous with a global shift in terrestrial vegetation distribution between 8 and 5 Myr ago, which has been interpreted by some studies as a floral response to decreasing partial pressures of carbon dioxide (p_{CO_2}) in the atmosphere^{4–6}. We infer a global decrease in carbon dioxide levels for this time interval that has not yet been identified in the sparse p_{CO_2} proxy record⁷ but is synchronous with global cooling and progressive glaciations^{8,9}.

Coccolithophores are unique among algae in that they use carbon both for calcification and for photosynthesis. Cultures of coccolithophores grown under ambient, CO_2 -limiting conditions show an unusually large array (up to 5‰) of non-equilibrium carbon and oxygen stable isotopic fractionations ($\delta^{13}\text{C}$ and $\delta^{18}\text{O}$)^{10,11}. These isotope 'vital effects', so called because they are thought to result from biological processes, are also evident in coccoliths from recent sediments and sediment traps. The isotopic difference between small and large coccoliths diminishes in cultures grown at elevated $[\text{CO}_2]$ (increased dissolved inorganic carbon concentration at constant pH)¹² (Fig. 1b) and is absent in fossil coccoliths from past Palaeocene greenhouse climates^{13,14}. We assert that vital effects reflect the adaptation of cellular carbon fluxes to aqueous CO_2 availability, and in a new model we reveal the origin of carbon isotope vital effects. We then evaluate the timing of the emergence of vital effects in the fossil record and its relationship to Cenozoic climate evolution and the long-term decrease in p_{CO_2} .

Photosynthesis in large cells may be more sensitive to limitation by diffusive CO_2 supply because of the lower ratio of surface area to volume (Supplementary Fig. 2). Active transport of HCO_3^- for photosynthesis is expected to be driven by the extent of diffusive CO_2 limitation, and may therefore differ between small and large cells. A new model (Supplementary Discussion) reveals the active HCO_3^- fluxes to the cell,

the site of photosynthesis (chloroplast) and the site of calcification (coccolith vesicle) required to explain the observed array of carbon isotopic fractionation into organic matter and coccolith calcite, ϵ_p and $\epsilon_{\text{coccolith}}$ respectively, observed in coccolithophore species of different sizes grown in culture at variable $[\text{CO}_2]$ (refs 12, 15) (Fig. 1). The model confirms that at low $[\text{CO}_2]$, active HCO_3^- transport to the chloroplast is increased at the expense of active HCO_3^- transport to the coccolith

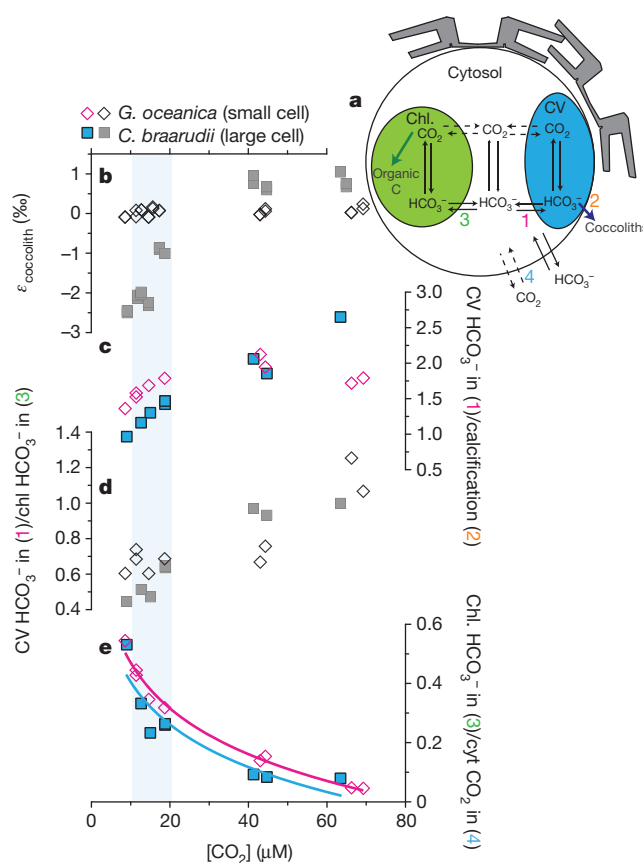
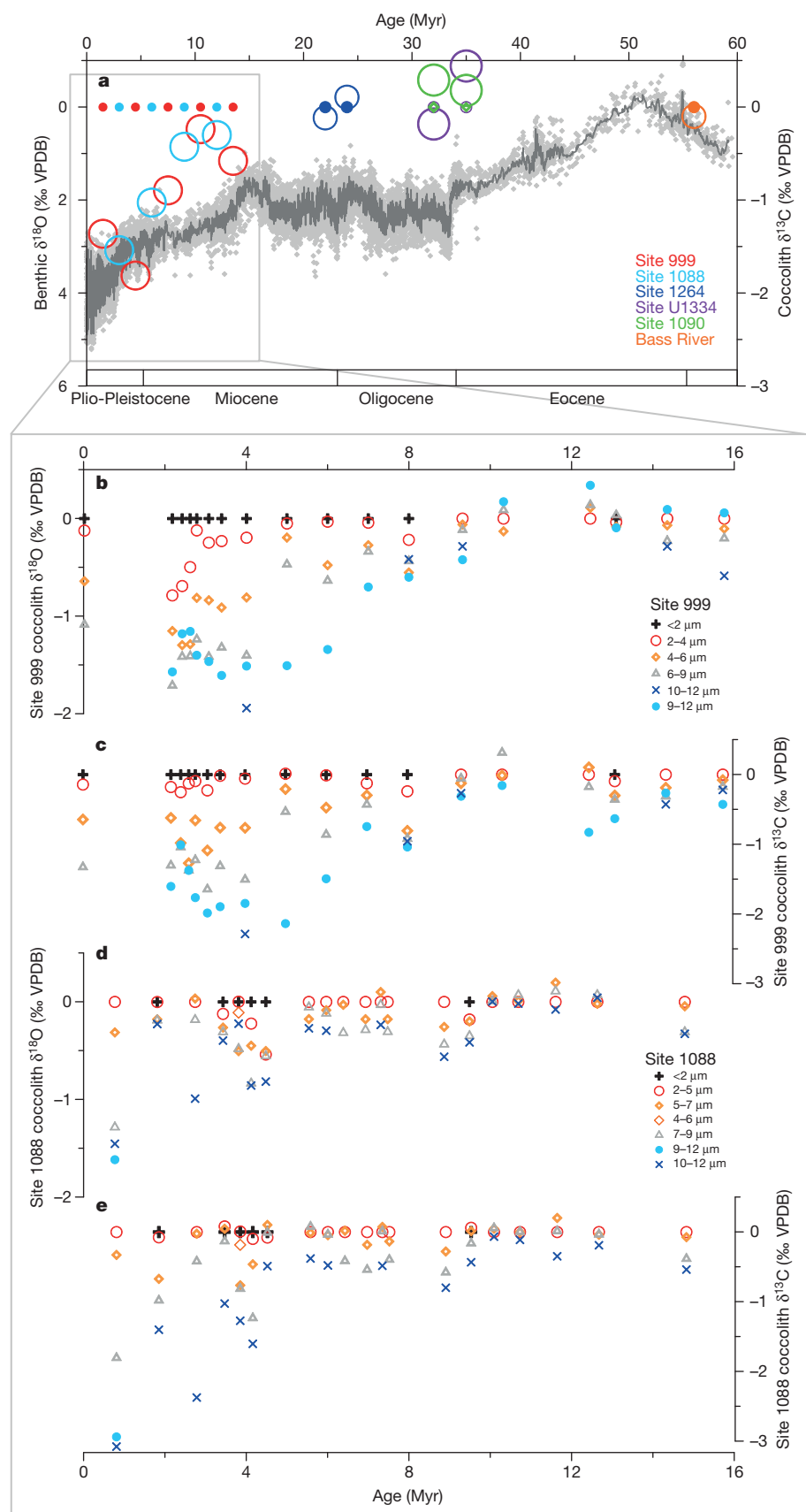


Figure 1 | HCO_3^- allocation to the chloroplast and coccolith vesicle inferred from $\epsilon_{\text{coccolith}}$ measured in culture. **a**, Simplified modelled coccolithophore carbon fluxes (details in Supplementary Fig. 1). CV, coccolith vesicle; Chl., chloroplast. Dashed black arrows represent passive fluxes, and solid black arrows represent active fluxes. **b**, $\epsilon_{\text{coccolith}}$ as a function of $[\text{CO}_2]$ (data from ref. 12; propagated analytical uncertainty 0.1‰). **c**, Coccolith vesicle HCO_3^- influx relative to calcification. **d**, Coccolith vesicle HCO_3^- influx relative to chloroplast HCO_3^- influx. **e**, Chloroplast HCO_3^- influx relative to diffusive CO_2 uptake by cell. Data in c–e are inferred from inverse model (Supplementary Information) using default parameters (Supplementary Table 1). Symbols in b–e: diamonds, *Gephyrocapsa oceanica*; squares, *Coccolithus pelagicus* subsp. *braarudii*. Blue shading indicates the range of steepest dependence of $\epsilon_{\text{coccolith}}$ on $[\text{CO}_2]$.

¹Geology Department, University of Oviedo, Jesus Arias de Velasco S/N, 33005, Oviedo, Asturias, Spain.

Figure 2 | Divergence of vital effects in coccoliths. **a**, Benthic foraminiferal $\delta^{18}\text{O}$ (ref. 9) (data points in light grey, smoothed with seven-point running mean) and $\delta^{13}\text{C}$ of smallest and largest coccoliths (coloured circles). All $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ values are measured against Vienna Pee Dee Belemnite (VPDB). See Supplementary Fig. 10 for complete size fraction data. Bubble size scales with approximate coccolith size. For the Neogene, mean values for 3-Myr time windows are shown from Sites 999 and 1088. The grey box denotes the time interval in **b–e** (16–0 Myr ago). **b**, **c**, $\delta^{18}\text{O}$ (**b**) and $\delta^{13}\text{C}$ (**c**) of different-sized coccoliths from Site 999. **d**, **e**, $\delta^{18}\text{O}$ (**d**) and $\delta^{13}\text{C}$ (**e**) of different-sized coccoliths from Site 1088. To remove secular trends and highlight differences between size fractions, all coccolith isotopes are normalized to the smallest coccolith size fraction in each sample. Note the different scales of $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ axes.



vesicle. A similar competitive reallocation of HCO_3^- to photosynthesis from calcification at low $[\text{CO}_2]$ has been shown in the laboratory¹⁶. As a consequence, at low $[\text{CO}_2]$ a smaller proportion of calcification is supported by a direct influx of HCO_3^- to the coccolith vesicle, decreasing

$\epsilon_{\text{coccolith}}$. This process is amplified in larger cells, which at low $[\text{CO}_2]$ feature the lowest proportion of calcification supported by direct influx of HCO_3^- to the coccolith vesicle. Consequently, the difference in $\epsilon_{\text{coccolith}}$ between large and small coccolithophores is greater at low

[CO₂]. Culture data and our model indicate that this relationship is nonlinear, with the steepest dependence of $\varepsilon_{\text{coccolith}}$ on [CO₂] over the range 12–19 μM (Fig. 1b). Vital effects in $\delta^{18}\text{O}$ have previously been ascribed to changes in the relative contributions of carbonate (CO₃²⁻) and HCO₃⁻ to coccolith calcite¹⁷, which produces an effect analogous to that generated by variable relative influx of CO₂ and HCO₃⁻ to the coccolith vesicle predicted by our $\delta^{13}\text{C}$ model (Supplementary Discussion).

Evaluation of $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ in size-separated coccoliths from five (Integrated) Ocean Drilling Program sites (Supplementary Methods and Supplementary Fig. 9) shows that vital effects of stable isotopes in coccoliths were minimal before and after the Eocene–Oligocene (about 34 Myr ago) and Oligocene–Miocene (about 23 Myr ago) transitions, and that large (more than 1‰) vital effects first appeared during the late Miocene to earliest Pliocene (about 7–5 Myr ago). A striking divergence in isotopic composition in different-sized coccoliths is demonstrated in records from two widely separated sites, Caribbean Site 999 and sub-Antarctic Site 1088 (Figs 2 and 3). In samples pre-dating 7 Myr ago, only small $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ differences (less than 0.75‰) between size fractions are observed. After the divergence, which begins at 6–7 Myr ago at Site 999 and 4–5 Myr ago at Site 1088, persistent vital effects of 1.5–3‰ in $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ are recorded, with large coccoliths consistently recording lighter $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ relative to smaller coccoliths (Fig. 2). We interpret this diachrony as a real lag that is too large to result from age model discrepancies (Supplementary Methods and Supplementary Fig. 11). We note that temporal changes in mean

coccolith size in the sediments do not affect our data from restricted coccolith size classes.

The marked increase in vital effects in coccoliths in the late Miocene cannot reflect an expansion into a wider range of depth habitats, because the $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ values in different-sized coccoliths are positively correlated (Fig. 2 and Supplementary Fig. 10), not negatively correlated as would be expected from depth segregation in the photic zone¹³. We also find no cause to suggest that the depth habitat of all coccolithophores at both sites migrated from deeper CO₂-enriched to shallower CO₂-depleted waters within the photic zone (Supplementary Discussion). At Site 999 it is possible that circulation changes associated with the gradual closure of the Central American Seaway about 14–3 Myr ago (ref. 18) stemmed the eastward flow of CO₂-rich upwelled water from the equatorial Pacific; however, the emergence of the Panama isthmus is not modelled to strongly affect circulation near Site 1088 (ref. 19). The shift to a large array of vital effects in coccoliths occurs at a time when there is no evidence for large changes in coccolithophore growth rate at either site, as indicated by coccolith Sr/Ca records (Supplementary Methods and Supplementary Fig. 5). A shift from predominantly (more than 70%) diagenetic calcite to primary coccolith calcite would be required to homogenize a 1.5‰ isotopic difference in primary $\delta^{18}\text{O}$ to the less than 0.6‰ recorded in older sediments (Supplementary Fig. 8). This is not consistent with the moderate to good coccolith preservation throughout the Miocene–Pliocene at both sites evident in scanning electron microscope images (Supplementary Figs 6 and 7), nor with Sr/Ca values, which confirm biogenic rather than abiogenic (diagenetic) Sr partitioning throughout the Miocene–Pliocene study interval (Supplementary Discussion). The presence of vital effects at the Pliocene end of both records, and their absence at the Miocene end, is unlikely to result from differences in species contributions in a given size fraction over time. Counts of coccoliths in all size fractions from end-member samples show that, despite changes in species composition and size distribution over the 16-Myr study interval, the genera or families dominating each size fraction remain similar (Supplementary Table 3). For example, at Site 1088, smallest and largest coccolith size fractions in both Pleistocene and Miocene end-member samples are dominated (more than 70% CaCO₃) by small reticulofenestrid and *Coccolithus pelagicus* coccoliths, respectively, yet only the Pleistocene sample records a large array (up to 3‰) of vital effects (Fig. 2).

Our model of coccolithophore carbon allocation suggests that the late Miocene emergence of vital effects represents a modification of carbon acquisition strategies of the cells as [CO₂] decreased below a critical threshold (Fig. 1). We propose that a decrease in p_{CO_2} caused tropical waters (Site 999) to fall below this [CO₂] threshold at about 7 Myr ago. Because CO₂ is more soluble in cold waters, a continued p_{CO_2} decline into the early Pliocene (about 5 Myr ago) was required before a similar limiting [CO₂] was reached in the cooler sub-Antarctic waters of Site 1088 (Supplementary Fig. 12).

The emergence of large-scale vital effects in coccoliths in the late Miocene, rather than at earlier transitions such as the Eocene–Oligocene or Oligocene–Miocene, for which important step decreases in p_{CO_2} are estimated from proxies and inferred from climate records^{20–23}, is consistent with culture data¹², which suggest low sensitivity of $\varepsilon_{\text{coccolith}}$ to [CO₂] variation above 19 μM . At typical concentrations of dissolved inorganic carbon in the surface ocean (2,050 μM) and estimated production temperatures for a typical mid-latitude site (20 °C; Supplementary Fig. 5), the range of maximum sensitivity (12–19 μM [CO₂]) corresponds to a p_{CO_2} in the range 575–375 parts per million by volume (p.p.m.v.). As [CO₂] decreases below 20 μM there is an exponential increase in the requirement for active HCO₃⁻ transport to the chloroplast (Supplementary Fig. 4). Since the late Miocene, further decreases in p_{CO_2} , even to low values typical of the last glacial¹³, have not resulted in a subsequent increase in the magnitude of size-related vital effects. One explanation could be that further decreases in [CO₂] were accompanied by a decrease in cellular calcification, thereby limiting further

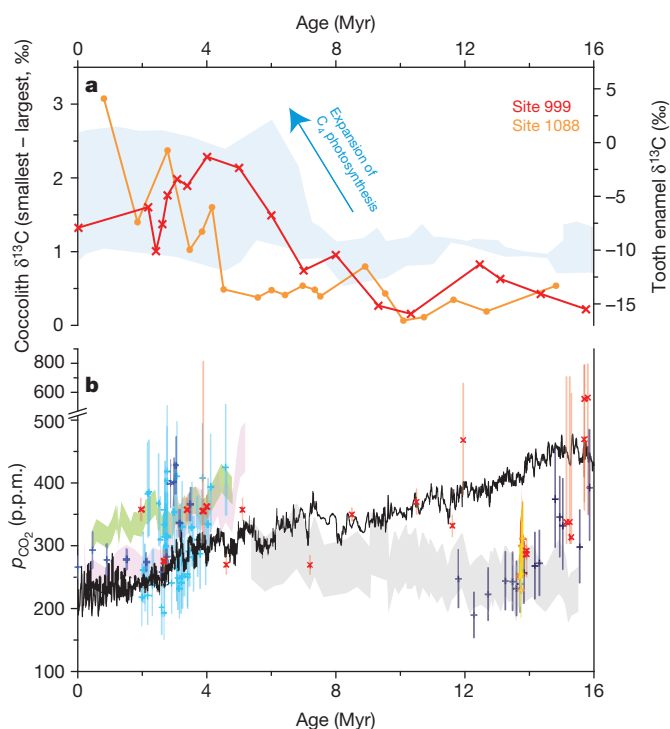


Figure 3 | Evolution of vital effects in coccoliths, C₄ photosynthesis and p_{CO_2} 16 Myr ago. a, $\delta^{13}\text{C}$ difference between smallest and largest coccolith size fractions at Sites 999 (red) and 1088 (orange) and the range of tooth enamel $\delta^{13}\text{C}$ values (blue shading; data from ref. 4; only North American data below 37° N are plotted; however, other regions show a similar pattern). The propagated analytical uncertainty on coccolith $\delta^{13}\text{C}$ differences is 0.07‰. **b**, Estimates of p_{CO_2} from various proxies: foraminifer boron isotopes (blue and yellow horizontal crosses), stomata (red diagonal crosses), alkenone $\delta^{13}\text{C}$ maximum and minimum estimates (pink, green, grey and orange shading), and inverse modelling of deep-sea $\delta^{18}\text{O}$ (black line). Note the change in scale at 500 p.p.m.v. Vertical error bars represent the uncertainty reported in published p_{CO_2} estimates. See Supplementary Information for p_{CO_2} data references and details of uncertainty derivation for each reference.

decreases in the supply of HCO_3^- to the coccolith vesicle relative to calcification. Decreased calcification in coccoliths of a given size over the Cenozoic could support the operation of such a mechanism^{24,25}.

Few p_{CO_2} proxy reconstructions cover the interval leading up to the divergence of vital effects in coccoliths (12–5 Myr ago). Alkenone-based records suggest low and stable p_{CO_2} during this interval (Fig. 3b). However, these estimates could be too low because of the nature of the applied corrections for temperature and phosphate concentrations^{22,26}. New alkenone-based p_{CO_2} estimates from the western tropical Atlantic covering the mid to late Miocene, although low in resolution, suggest substantially higher values (400–500 p.p.m.v.)²⁷. Although uncertainties remain large, stomatal proxies indicate a p_{CO_2} decrease⁷, consistent with inverse modelling of climate data⁸ (Fig. 3b). Our data suggest that substantial surface ocean cooling over the past 15 Myr, up to 14 °C in the subtropics²⁸, may reflect an important global p_{CO_2} decrease that is poorly resolved by existing p_{CO_2} proxy records, rather than a decoupling of atmospheric CO_2 forcing and climate as suggested by some authors²⁸.

The appearance of large-scale vital effects in coccoliths between 7 and 5 Myr ago is synchronous with a global expansion in terrestrial C_4 plants (that is, those using the C_4 photosynthetic pathway; mostly tropical grasses) relative to C_3 plants (primarily trees) in low latitudes and mid-latitudes^{4–6,29} (Fig. 3a). In some regions, such as the Himalayan foreland and Arabian Peninsula, it has been suggested that a shift to increasingly arid conditions was the dominant driver of the late Miocene rise in C_4 plants²⁹. However, the shift to C_4 dominance has also been widely interpreted as a response to decreasing p_{CO_2} , because at low ratios of atmospheric CO_2 to O_2 concentrations C_4 plants have a competitive advantage over C_3 plants^{4–6}. The presence of a biochemical carbon-concentrating mechanism allows C_4 plants to decrease energetically costly photorespiration rates, and also to decrease stomatal conductance (a measure of the rate at which water and CO_2 can diffuse in or out of the leaf), thus decreasing water loss. Conditions that favour C_4 over C_3 plants are suggested to occur below a p_{CO_2} of about 500 p.p.m.v. when accompanied by high temperatures during the growing season (that is, at low latitudes), or at lower p_{CO_2} in cooler climates^{4,5}. Thus, both terrestrial and marine photosynthesizers may be showing adaptation at a common p_{CO_2} threshold.

We show that the large array of isotopic fractionations in modern coccolith carbonate is indicative of the operation of strong carbon-concentrating mechanisms in coccolithophore cells, which became highly significant since the latest Miocene. We speculate that this change occurred as a threshold response to increased CO_2 limitation, beginning in the late Miocene in the tropical oceans and progressing to higher latitudes by the earliest Pliocene. This increase in the degree of active carbon uptake by coccolithophores will need to be accounted for in the application of ϵ_p to estimates of $[\text{CO}_2]$ (ref. 30). The relatively low $[\text{CO}_2]$ threshold suggested to have driven the late Miocene diversification of coccolithophore carbon acquisition strategies is consistent with estimates of less than 500 p.p.m.v. p_{CO_2} required to promote the tropical C_4 -dominated ecosystems that also expanded over this interval^{4–6}. We speculate that such a low p_{CO_2} threshold, affecting both marine and terrestrial primary producers, could be reversed within decades as a result of rapid anthropogenic CO_2 release and absorption by the ocean.

METHODS SUMMARY

We adapt a model for the $\delta^{13}\text{C}$ composition of photosynthetically fixed carbon in diatoms³¹ with an additional module for the coccolith vesicle, allowing us to simulate the $\delta^{13}\text{C}$ of coccolith calcite as a function of the passive and active carbon fluxes into the coccolith vesicle and cell (model ACTI-CO; see Supplementary Discussion). Coccolith size fractions were separated from bulk IODP sediment samples by using site-specific and interval-specific settling and microfiltration protocols (Supplementary Methods). Coccolith $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ were measured on a Nu Perspective dual-inlet isotope ratio mass spectrometer connected to a NuCarb carbonate preparation system, with an analytical precision of 0.06‰ for $\delta^{18}\text{O}$ and 0.05‰ for $\delta^{13}\text{C}$ (1 σ), at Oviedo University. Mean reproducibility, based on duplicate analyses of splits of 21 random samples from Sites 999 and 1088, is

0.08‰ for $\delta^{18}\text{O}$ and 0.06‰ for $\delta^{13}\text{C}$ (1 σ). Sr/Ca was determined in two coccolith size fractions at both Sites 999 and 1088. Reducing and ion-exchange treatments were first applied to clean the samples, followed by gentle dissolution in acetic acid with an ammonium acetate buffer for 12 h. Calcium content was measured on a split of all samples, which were then diluted to constant calcium concentrations for Sr/Ca analysis by inductively coupled plasma optical emission spectroscopy on a Thermo ICAP DUO 6300 at Oviedo University. Sr/Ca data were corrected for site-specific variations in sea surface temperature (Supplementary Methods). All coccolith counts were performed on standard smear slides with a light microscope under cross-polarized light at a magnification of $\times 1,250$. To assess preservation, coccolith samples on polycarbonate filters were mounted onto a stub, coated with gold and imaged on a Jeol 6610LV scanning electron microscope.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 January; accepted 12 July 2013.

1. Tortell, P. D. Evolutionary and ecological perspectives on carbon acquisition in phytoplankton. *Limnol. Oceanogr.* **45**, 744–750 (2000).
2. Raven, J. A., Giordano, M., Beardall, J. & Maberly, S. C. Algal and aquatic plant carbon concentrating mechanisms in relation to environmental change. *Photosynth. Res.* **109**, 281–296 (2011).
3. Reinfelder, J. R. Carbon concentrating mechanisms in eukaryotic marine phytoplankton. *Annu. Rev. Mar. Sci.* **3**, 291–315 (2011).
4. Cerling, T. E. *et al.* Global vegetation change through the Miocene–Pliocene boundary. *Nature* **389**, 153–158 (1997).
5. Ehleringer, J. R., Cerling, T. E. & Helliker, B. R. C_4 photosynthesis, atmospheric CO_2 , and climate. *Oecologia* **112**, 285–299 (1997).
6. Zhang, C. *et al.* C_4 expansion in central Inner Mongolia during the latest Miocene and early Pliocene. *Earth Planet. Sci. Lett.* **287**, 311–319 (2009).
7. Beerling, D. J. & Royer, D. L. Convergent Cenozoic CO_2 history. *Nat. Geosci.* **4**, 418–420 (2011).
8. van de Wal, R. S., de Boer, B., Lourens, L. J., Köhler, P. & Bintanja, R. Reconstruction of a continuous high-resolution CO_2 record over the past 20 million years. *Clim. Past* **7**, 1459–1469 (2011).
9. Zachos, J. C., Dickens, G. R. & Zeebe, R. E. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* **451**, 279–283 (2008).
10. Dudley, W. C., Blackwelder, P., Brand, L. & Duplessy, J. C. Stable isotopic composition of coccoliths. *Mar. Micropaleontol.* **10**, 1–8 (1986).
11. Ziveri, P. *et al.* Stable isotope ‘vital effects’ in coccolith. *Earth Planet. Sci. Lett.* **210**, 137–149 (2003).
12. Rickaby, R. E. M., Henderiks, J. & Young, J. N. Perturbing phytoplankton: response and isotopic fractionation with changing carbonate chemistry in two coccolithophore species. *Clim. Past* **6**, 771–785 (2010).
13. Bolton, C. T., Stoll, H. M. & Mendez Vicente, A. Vital effects in coccolith calcite: Cenozoic climate- p_{CO_2} drove the diversity of carbon acquisition strategies in coccolithophores? *Paleoceanography* **27**, PA4204 (2012).
14. Stoll, H. M. Limited range of interspecific vital effects in coccolith stable isotopic records during the Paleocene–Eocene thermal maximum. *Paleoceanography* **20**, PA1007 (2005).
15. Rost, B., Zondervan, I. & Riebesell, U. Light-dependent carbon isotope fractionation in the coccolithophorid *Emiliania huxleyi*. *Limnol. Oceanogr.* **47**, 120–128 (2002).
16. Bach, L. T. *et al.* Dissecting the impact of CO_2 and pH on the mechanisms of photosynthesis and calcification in the coccolithophore *Emiliania huxleyi*. *New Phytol.* **199**, 121–134 (2013).
17. Ziveri, P., Thoms, S., Probert, I., Geisen, M. & Langer, G. A universal carbonate ion effect on stable isotope ratios in unicellular planktonic calcifying organisms. *Biogeosciences* **9**, 1025–1032 (2012).
18. Kameo, K. & Sato, T. Biogeography of Neogene calcareous nanofossils in the Caribbean and the eastern equatorial Pacific—floral response to the emergence of the Isthmus of Panama. *Mar. Micropaleontol.* **39**, 201–218 (2000).
19. Schneider, B. & Schmittner, A. Simulating the impact of the Panamanian seaway closure on ocean circulation, marine productivity and nutrient cycling. *Earth Planet. Sci. Lett.* **246**, 367–380 (2006).
20. Kürschner, W. M., Kvacek, Z. & Dilcher, D. L. The impact of Miocene atmospheric carbon dioxide fluctuations on climate and the evolution of terrestrial ecosystems. *Proc. Natl Acad. Sci. USA* **105**, 449–453 (2008).
21. Pagani, M., Arthur, M. A. & Freeman, K. H. Miocene evolution of atmospheric carbon dioxide. *Paleoceanography* **14**, 273–292 (1999).
22. Pagani, M. *et al.* The role of carbon dioxide during the onset of Antarctic glaciation. *Science* **334**, 1261–1264 (2011).
23. Pearson, P. N., Foster, G. L. & Wade, B. S. Atmospheric carbon dioxide through the Eocene–Oligocene climate transition. *Nature* **461**, 1110–1113 (2009).
24. Preiss-Daimler, I., Baumann, K.-H. & Heinrich, R. Carbonate budget mass estimates for Neogene disconformities from the Equatorial Atlantic (Ceara Rise: ODP Site 927). *J. Micropaleontol.* **31**, 169–178 (2012).
25. Stoll, H. M. *et al.* Insights on coccolith chemistry from a new ion probe method for analysis of individually picked coccoliths. *Geochem. Geophys. Geosyst.* **8**, Q06020 (2007).

26. Pagani, M., Liu, Z., LaRiviere, J. & Ravelo, A. C. High Earth-system climate sensitivity determined from Pliocene carbon dioxide concentrations. *Nat. Geosci.* **3**, 27–30 (2009).
27. Zhang, Y. G., Pagani, M., Liu, Z., Bohaty, S. M. & DeConto, R. A 40 million-year history of atmospheric CO₂. *Phil. Trans. R. Soc. A*, <http://dx.doi.org/10.1098/rsta.2013.0096> (in the press).
28. LaRiviere, J. P. *et al.* Late Miocene decoupling of oceanic warmth and atmospheric carbon dioxide forcing. *Nature* **486**, 97–100 (2012).
29. Huang, Y., Clemens, S. C., Liu, W., Wang, Y. & Prell, W. L. Large-scale hydrological change drove the late Miocene C₄ plant expansion in the Himalayan foreland and Arabian Peninsula. *Geology* **35**, 531–534 (2007).
30. Laws, E. A., Popp, B. N., Cassar, N. & Tanimoto, J. ¹³C discrimination patterns in oceanic phytoplankton: likely influence of CO₂ concentrating mechanisms, and implications for palaeoreconstructions. *Funct. Plant Biol.* **29**, 323–333 (2002).
31. Hopkinson, B. M., Dupont, C. L., Allen, A. E. & Morel, F. M. M. Efficiency of the CO₂-concentrating mechanism of diatoms. *Proc. Natl Acad. Sci. USA* **108**, 3830–3837 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank L. Abrevaya and A. Mendez for laboratory assistance, and K. Lawrence for access to unpublished data. This work used samples provided by the (Integrated) Ocean Drilling Program (IODP). The IODP is sponsored by the US National Science Foundation and participating countries under management of the IODP Management International, Inc. Funding for this research was provided by European Research Council grant UE-09-ERC-2009-STG-240222-PACE (H.M.S.) and a DuPont Young Professor Award to H.M.S.

Author Contributions C.T.B. and H.M.S. designed the study and wrote the paper. C.T.B. separated coccoliths and performed stable isotope, light microscope and scanning electron microscope analyses. H.M.S. designed and ran the model.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.T.B. (cbolton@geol.uniovi.es) or H.M.S. (hstoll@geol.uniovi.es).

Rapid, climate-driven changes in outlet glaciers on the Pacific coast of East Antarctica

B. W. J. Miles¹, C. R. Stokes¹, A. Vieli^{1†} & N. J. Cox¹

Observations of ocean-terminating outlet glaciers in Greenland and West Antarctica^{1–6} indicate that their contribution to sea level is accelerating as a result of increased velocity, thinning and retreat^{7–11}. Thinning has also been reported along the margin of the much larger East Antarctic ice sheet¹, but whether glaciers are advancing or retreating there is largely unknown, and there has been no attempt to place such changes in the context of localized mass loss^{7,9} or climatic or oceanic forcing. Here we present multidecadal trends in the terminus position of 175 ocean-terminating outlet glaciers along 5,400 kilometres of the margin of the East Antarctic ice sheet, and reveal widespread and synchronous changes. Despite large fluctuations between glaciers—linked to their size—three epochal patterns emerged: 63 per cent of glaciers retreated from 1974 to 1990, 72 per cent advanced from 1990 to 2000, and 58 per cent advanced from 2000 to 2010. These trends were most pronounced along the warmer western South Pacific coast, whereas glaciers along the cooler Ross Sea coast experienced no significant changes. We find that glacier change along the Pacific coast is consistent with a rapid and coherent response to air temperature and sea-ice trends, linked through the dominant mode of atmospheric variability (the Southern Annular Mode). We conclude that parts of the world's largest ice sheet may be more vulnerable to external forcing than recognized previously.

Ice sheets lose mass through melting (surface or basal) and dynamic changes (for example, the acceleration and retreat of outlet glaciers). For the Greenland ice sheet, these two components have made approximately equal contributions to its recent negative mass balance¹². In Antarctica, surface melt is much less important, but the West Antarctic ice sheet is thought to be vulnerable to oceanic warming because large parts of its bed lie below sea level¹³. Recent estimates have also confirmed its negative mass balance^{9–11}. In contrast, the mass balance of the much larger East Antarctic ice sheet (EAIS) is closer to equilibrium or slightly positive^{7,9–11}, but recent thinning (2003–2007) has been detected on several major outlet glaciers¹, resulting in negative imbalances in some catchments^{7,9}. Similar thinning of outlet glaciers in the Greenland and West Antarctic ice sheets and the Antarctic Peninsula over the past two decades has been associated with glacier acceleration and changes at their termini^{2,6,8,14}, predominantly retreat and the thinning or collapse of ice shelves^{5,6}. However, unlike for the Greenland^{2,15} and West Antarctic⁴ ice sheets and the Antarctic Peninsula¹⁶, there has been no comprehensive analysis of glacier terminus positions in East Antarctica. Measurements on a small number of glaciers (<20) revealed cyclic behaviour with no obvious trend or a reduction in their floating area since the 1950s^{17,18}.

Here we use approximately 300 satellite images (spanning 1963–2012) to map the terminus position of a comprehensive set of 175 glaciers along 5,400 km of the EAIS, stretching from Queen Mary Land (longitude 90° E) to Victoria Land (170° E) (Methods Summary and Supplementary Table 1). This region represents about a third of the EAIS margin and was selected because it encompasses two regions of mass loss (Wilkes Land and northern Victoria Land)^{7,9}; because large parts are grounded below sea level¹⁹, which may enhance its vulnerability to oceanic forcing; and because the absence of large ice shelves makes

individual glacier termini readily identifiable. Glaciers in this region encompass a range of widths (0.65 to ~57 km) and flow speeds (~155 to ~1,400 m yr⁻¹) and all calve into the ocean, with most (~90%) possessing floating extensions and many (~84%) unconstrained by lateral boundaries (for example fjord walls) at their termini. To minimize the influence of short-term interannual variations and major (potentially stochastic) calving events that are known to occur^{17,18}, we focused on large numbers of glaciers at approximately decadal time steps (1974, 1990, 2000 and 2010), but the measurement years were dictated by the availability of imagery when most glaciers ($n > 130$) could be mapped.

A small set of glaciers ($n = 38$) measured in 1963 and 2010 show an overall pattern of retreat (median terminus position change, -12.9 m yr⁻¹; mean, -61.2 m yr⁻¹) (Table 1). However, a larger set ($n = 132$) measured in 1974 and 2010, show very little overall change (median, 0.7 m yr⁻¹; mean, -2.7 m yr⁻¹), but there are clear phases of advance and retreat within this period (Fig. 1). Specifically, 1974–1990 was characterized by retreat (63% of glaciers) at a median rate of -12.5 m yr⁻¹ (mean, -43.3 m yr⁻¹). From 1990 to 2000, however, this trend was reversed, when 72% of glaciers advanced at a median rate of 19.7 m yr⁻¹ (mean, 43.1 m yr⁻¹). During the most recent period, 2000–2010, the number of advancing glaciers fell to 58% and the median decreased to 8.4 m yr⁻¹ (mean, -17.9 m yr⁻¹).

The magnitude of advance or retreat experienced by different glaciers varies considerably (Fig. 2) and is linked to their width, which is correlated with glacier speed (Supplementary Fig. 1). Thus, large glaciers with higher flow speeds tend to undergo the largest changes, in both advance and retreat phases. Several large glaciers (for example those >15 km wide) experienced major calving events that caused retreats of tens of kilometres followed by re-advance, indicating cyclic behaviour potentially unrelated to external forcing (Supplementary Figs 2 and 3). This process might introduce considerable variability and obscure any trends. However, the inclusion or exclusion of large glaciers seems to have little effect: the switch from retreat (1974–1990) to advance (1990–2000) is very highly significant ($P < 0.0005$), irrespective of whether large glaciers are included (Fig. 2a, b and Supplementary Table 2). Significant differences ($P < 0.05$) are also found between the second and third periods (or epochs) we consider, owing to an increased number of glaciers undergoing retreat from 2000–2010 compared with 1990–2000. However, the significance levels are lower because of the more even mix of advance and retreat in the most recent period (Supplementary Table 3).

These trends in terminus retreat and advance are most pronounced along the western South Pacific coast (Fig. 2c), where the changes from retreat (1974–1990) to advance (1990–2000) and back to retreat (2000–2010) are significant (Supplementary Tables 2 and 3). In contrast, those glaciers facing the Ross Sea (Fig. 2d) show no significant differences between any epochal divisions. Thus, there is a regional difference between Pacific-facing glaciers, which generally lie along the Antarctic circle (66° 33' S), and those farther south that face the Ross Sea (Fig. 1), suggesting a potential link to climate forcing.

¹Department of Geography, Durham University, Science Site, South Road, Durham DH1 3LE, UK. †Present address: Department of Geography, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland.

Table 1 | Changes in glacier terminus position during different epochs from 1963 to 2010

	1963–2010	1974–2010	1974–1990	1990–2000	2000–2010
Number of glaciers	38	132	131 (128)	168 (128)	171 (128)
Advanced (%)	32	53	37 (36)	72 (72)	58 (63)
Retreated (%)	68	47	63 (64)	28 (28)	42 (37)
Mean terminus change (m yr^{-1})	−61.2	−2.7	−43.3 (−44.9)	43.1 (30.9)	−17.9 (30.6)
Median terminus change (m yr^{-1})	−12.9	0.7	−12.5 (−12.8)	19.7 (14.5)	8.4 (13.7)

Negative values denote retreat and positive values denote advance. Data not in parentheses include all glacier measurements available at each time step (Fig. 1). For comparability, values in parentheses are for 128 glaciers measured at every time step, which reveal near-identical trends. Mean values are sensitive to extreme events (that is, calving of major ice tongues), suggesting that median values are more robust.

Mean annual and mean winter (June–August) air temperatures at three stations along the Pacific coast (Dumont d'Urville, Casey and Mirny; Fig. 1) are about 9 and 12 °C warmer, respectively, than at Scott Station along the Ross Sea coast (Supplementary Fig. 4). Annual and winter data show no clear trends at any station. However, the 1974–1990 mean austral summer temperature was significantly warmer ($\sim 1^\circ\text{C}$) than the 1990–2000 mean at all Pacific stations, but not at Scott station (Fig. 3 and Supplementary Table 4). The long-term warming before 1990 is most pronounced at Casey (0.22 °C per decade) and Dumont d'Urville (0.15 °C per decade), where daily summer temperatures are close to and occasionally greater than 0 °C, but the Ross Sea trend shows a cooling over the same period (1963–1990). The most recent period (2000–2010) was slightly warmer ($\sim 0.5^\circ\text{C}$) than the 1990–2000 period at all four stations, but the differences are not statistically significant and temperatures were not as high as in the 1974–1990 epoch.

Significant changes in air temperature along the Pacific coastline are therefore consistent with significant changes in terminus position in that region (Fig. 3), with the relatively warm period in the 1970s and 1980s associated with glacier retreat, and subsequent cooling during the 1990s coinciding with advance. Warming in the first half of the 2000s might also explain the larger number of glaciers that retreated in the period 2000–2010, compared with 1990–2000 (Fig. 3), but both warming and retreat are shorter term than before 1990, and since 2005 there has been a return to cooling. This is consistent with the larger range in terminus position change and the weaker, but still significant, differences. In contrast, there are no significant trends in air temperatures or glacier behaviour in the colder Ross Sea region.

These patterns hint that the response of outlet glaciers along the Pacific Coast may be related to the degree of surface melting. Indeed,

meltwater ponds are identifiable on glaciers along this coast (Supplementary Fig. 2), and January temperatures at Casey (1974–1990) were, on average, only cooler by 0.7 °C than at Faraday (Antarctic Peninsula), where glacier retreat has been linked to atmospheric warming¹⁶. Increased surface melt during warmer-than-average summers has the potential to enhance the opening of crevasses close to the glacier terminus, and, hence, calving, through hydraulic fracturing²⁰, as suggested for ice-shelf break-up in the Antarctic Peninsula²¹. This may partly explain the relationship between austral summer temperatures and terminus change, and this hypothesis is supported by the lack of significant trends for glaciers located in the much colder Ross Sea region (Figs 2 and 3).

Although it is appealing to invoke this relationship between terminus change and air temperatures, it is unlikely that air temperature is the only or most important forcing. Trends in air temperature are connected to synchronous changes in the ocean–atmosphere system through the dominant mode of atmospheric variability, namely the Southern Annular Mode (SAM), which influences wind speed and direction, sea-ice concentrations, sea surface temperatures and coastal ocean upwelling^{22–26}. Positive phases of the SAM index, increasingly common during the last two decades (Fig. 3), and linked to both increased greenhouse gas concentrations and ozone depletion^{23,26}, are associated with cooler temperatures over East Antarctica, increased sea-ice concentrations and enhanced westerly airflows^{22,23,25}. Above-average fast-ice extent along the Pacific coast has also been noted in the study region in the 1990s and has been related to a change in wind direction from predominantly offshore to more along shore²⁴. Indeed, several studies report increasing sea-ice concentrations in the study region from approximately 1980 to 2010^{22,23,25,27,28}, which is consistent with the predominance of glacier advance since 1990, when above-average

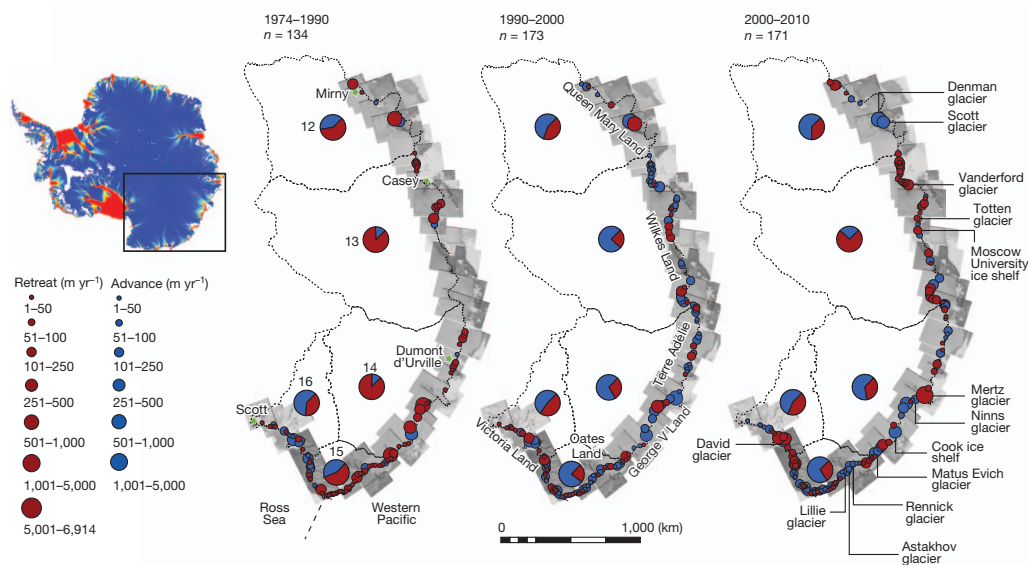


Figure 1 | Spatial and temporal variations in EAIS glacier terminus position from all measurements in 1974, 1990, 2000 and 2010. The rate of terminus position change for each glacier and period is shown by the single-colour circles (see key for sign and magnitude). Pie charts show the percentage of glaciers advancing and retreating in each major drainage basin (DB 12–16 (as marked

beside pie charts), from, for example, refs 9, 11). Climate stations referred to in this study (Fig. 3) are indicated by stars and the location map (top left) shows surface flow speed over Antarctica³⁰ with fast flow zones (for example $>500 \text{ m yr}^{-1}$) in red to yellow. Grey squares show coverage of individual Landsat scenes (180 km \times 180 km).

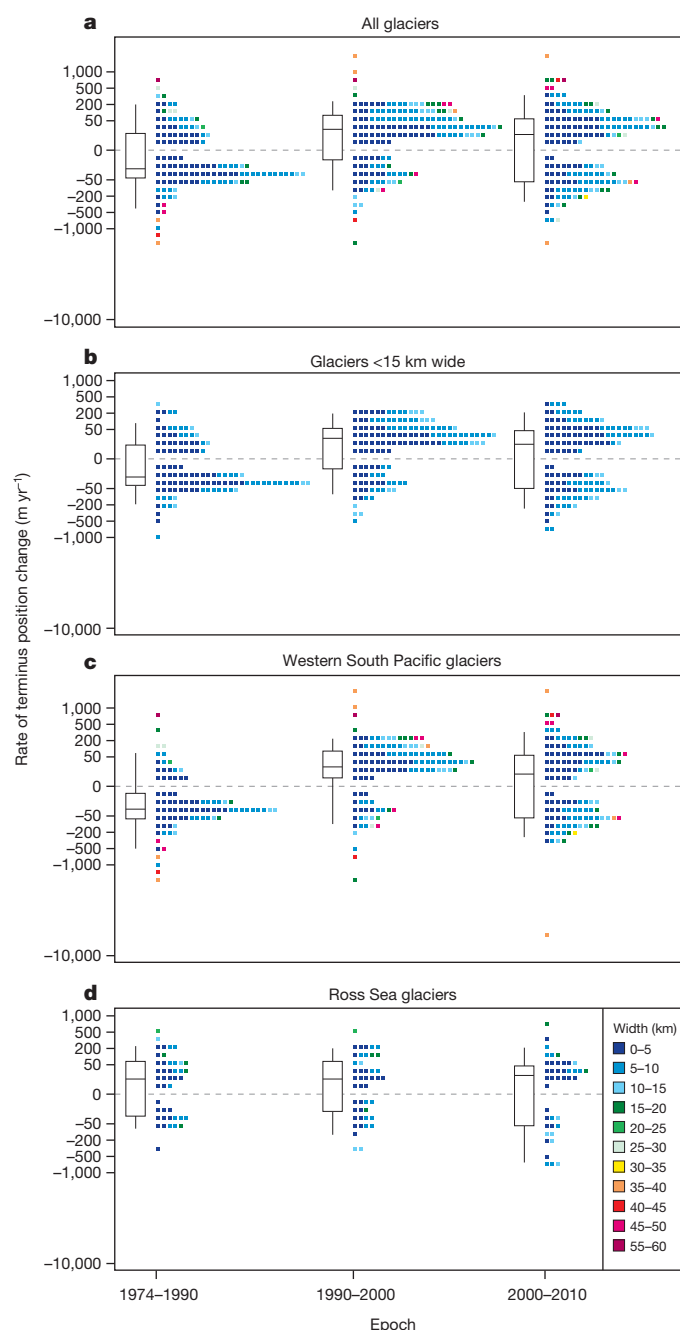


Figure 2 | Changes in glacier terminus position for each epoch for different sets of glaciers. Data for all glacier measurements are shown (a), alongside subsamples of glaciers <15 km wide (b), those facing the western South Pacific (c) and those facing the Ross Sea (d) (Fig. 1). Glacier data are shaded by width (see key), and box-and-whisker plots show the median (horizontal line), the 25th and 75th percentiles (box), and the 5th and 95th percentiles (whisker ends) on a cube-root scale (y axis). Significant differences between the 1974–1990 and 1990–2000 epochs and the 1990–2000 and 2000–2010 epochs are found for all samples of glaciers (a–c), apart from those facing the Ross Sea (d) (Supplementary Tables 2 and 3).

sea-ice and fast-ice concentrations could have suppressed calving by increasing back-pressure on glacier termini²⁹. In contrast, reduced sea-ice concentrations from the 1950s to the mid 1970s²⁷ are consistent with glacier retreat during the 1960s and 1970s, when air temperatures were also increasing along the Pacific coast (Fig. 3).

A complicating factor is that positive phases of the SAM are associated with increased coastal upwelling of warmer Circumpolar Deep Water^{22,25}. Intrusion of this water onto the continental shelf could

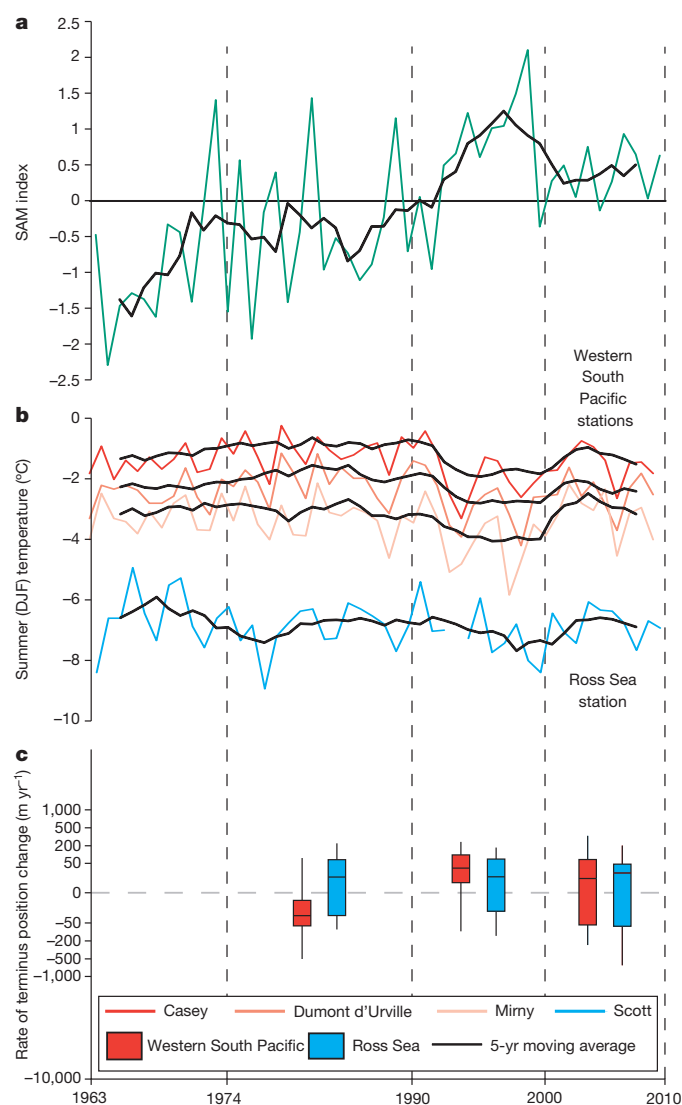


Figure 3 | Time series of the SAM and summer air temperature data alongside changes in glacier terminus positions. The December–May SAM index (a) and mean summer air temperature trends from the three Pacific Stations and one Ross Sea station (b) (Fig. 1) are shown alongside corresponding changes in glacier terminus position (c). Box-and-whisker plots show the median (horizontal line), the 25th and 75th percentiles (box), and the 5th and 95th percentiles (whisker ends) on a cube-root scale (y axis). Mean summer temperatures are calculated from mean monthly values of December, January and February; that is, data for 1974 are from December 1974, January 1975 and February 1975.

result in increased basal melting and weakening of ice tongues or shelves^{14,18}, but there are few deep submarine troughs within the study area¹⁴, and it would seem from our data that this process is yet to exert a major influence. Rather, despite the limitations imposed by our decadal measurements, we highlight tentative correlations between terminus position change and both air temperatures and the SAM index, which suggests that a change in mean summer temperature by $\sim 1^\circ\text{C}$ is manifest as a median terminus position change of 0.5 km per decade (Supplementary Figs 5 and 6).

Finally, glacier thinning has recently occurred along the Pacific coast in Wilkes Land (DB 12/13) and Victoria Land^{1,9} (DB 14/15). Elsewhere in Antarctica, similar rates of thinning have been linked to retreat and a reduction in buttressing, causing flow acceleration^{4–6}, but similar accelerations have not been reported in our study area⁷. The region of most pronounced mass loss^{1,7,9} (DB 13) is the only drainage basin to show a significant return to retreat from 2000 to 2010 (Supplementary

Table 6), but a wider comparison indicates that although glaciers that are thickening exhibit very little terminus change, those that are thinning are associated with both retreat and substantial advance (Supplementary Fig. 7). This indicates a more complex coupling between glacier discharge (for example velocity and elevation change) and terminus position than has been observed in the Greenland and West Antarctic ice sheets and the Antarctic Peninsula, because the floating extensions of most glacier termini in our study area are unconstrained and they do not exert any substantial buttressing. It may be that any future warming, perhaps driven by oceanic warming, or ozone recovery that results in a more negative SAM index^{23,26}, will thin or remove unconstrained ice tongues or shelves along the Pacific coast, such that terminus retreat has greater potential to induce dynamic mass loss, as observed elsewhere around Antarctica^{5,14}. However, although we detect a previously unrecognized widespread, rapid and synchronous response to large-scale atmospheric and oceanic variability, there is a clear requirement to understand the precise drivers of glacier dynamics to interpret and predict near-future mass loss from the EAIS. In particular, our results imply that the vulnerability of large parts of the EAIS margin requires urgent reassessment.

METHODS SUMMARY

We used optical satellite imagery to map the terminus position of 175 neighbouring outlet glaciers. To remove intra- and interannual variability on short timescales, termini were mapped at the end of the austral summer at approximately decadal time steps. Measurements were made on approximately 300 Landsat satellite images. The requirement for a comprehensive sample of glaciers spanning 5,400 km of coastline meant that only four main time steps allowed cloud-free mapping of the majority of glacier termini in the study area: 1974, 1990, 2000 and 2010 (Supplementary Table 1). A subsample of glaciers (38) were mapped with Argon imagery from 1963, but few of these could be re-measured in 1974 and they were widely spread geographically, which is why we exclude this epoch from detailed analysis. The accuracy of the mapping was dictated by co-registration of imagery and is ± 75 to ± 210 m for Landsat imagery and up to ± 420 m for some Argon imagery. Overall, 85% of measurements have an error below ± 180 m, which is comparable to a study from the Antarctic Peninsula¹⁶ and sufficient for extracting decadal trends. To account for uneven changes along the calving front, termini were digitized within a reference box that delineated the sides of the glacier². The mean retreat distance was calculated as the area change between each measurement, divided by glacier width. For comparison with elevation change measurements¹ (2003–2007), we mapped a sample of glaciers in austral summer 2006–2007 (Supplementary Fig. 7). Monthly mean surface air temperatures (Fig. 1) were extracted from the SCAR Met READER project (<http://www.antarctica.ac.uk/met/READER/>), and monthly values of the SAM index were obtained from <http://www.nerc-bas.ac.uk/icd/gima/sam.html>. To test for significance differences in terminus change and air temperatures between epochs, we used Student's *t*-test, the Wilcoxon signed-rank test and the Wilcoxon rank-sum test, where appropriate.

Full Methods and any associated references are available in the online version of the paper.

Received 8 November 2012; accepted 14 June 2013.

1. Pritchard, H. D., Arthern, R. J., Vaughan, D. G. & Edwards, L. A. Extensive dynamic thinning on the margins of the Greenland and Antarctic ice sheets. *Nature* **461**, 971–975 (2009).
2. Moon, T. & Joughin, I. Changes in ice front position on Greenland's outlet glaciers from 1992 to 2007. *J. Geophys. Res.* **113**, F02022 (2008).
3. Wingham, D., Wallis, D. W. & Shepherd, A. Spatial and temporal evolution of Pine Island Glacier thinning, 1995–2006. *Geophys. Res. Lett.* **36**, L17501 (2009).
4. Macgregor, J. A., Catania, G. A., Markowski, M. S. & Andrews, A. G. Widespread rifting and retreat of ice-shelf margins in the eastern Amundsen Sea Embayment between 1972 and 2011. *J. Glaciol.* **58**, 458–466 (2012).

5. Rignot, E. *et al.* Accelerated ice discharge from the Antarctic Peninsula following the collapse of Larsen B ice shelf. *Geophys. Res. Lett.* **31**, L18401 (2004).
6. Pritchard, H. D. & Vaughan, D. G. Widespread acceleration of tidewater glaciers on the Antarctic Peninsula. *J. Geophys. Res.* **112**, F03S29 (2007).
7. Rignot, E. *et al.* Recent Antarctic mass loss from radar interferometry and regional climate modelling. *Nature Geosci.* **1**, 106–110 (2008).
8. Nick, F. M., Vieli, A., Howat, I. M. & Joughin, I. Large-scale changes in Greenland outlet glacier dynamics triggered at the terminus. *Nature Geosci.* **2**, 110–114 (2009).
9. King, M. A. *et al.* Lower satellite-gravimetry estimates of Antarctic sea-level contribution. *Nature* **491**, 586–589 (2012).
10. Shepherd, A. *et al.* A reconciled estimate of ice-sheet mass balance. *Science* **338**, 1183–1189 (2012).
11. Zwally, H. J. & Giovinetto, M. B. Overview and assessment of Antarctic ice sheet mass balance estimates: 1992–2009. *Surv. Geophys.* **32**, 351–376 (2011).
12. van den Broeke, M. R. *et al.* Partitioning recent Greenland mass loss. *Science* **326**, 984–986 (2009).
13. Joughin, I. & Alley, R. B. Stability of the West Antarctic ice sheet in a warming world. *Nature Geosci.* **4**, 506–513 (2011).
14. Pritchard, H. D. *et al.* Antarctic ice-sheet loss driven by basal melting of ice shelves. *Nature* **484**, 502–505 (2012).
15. Björk, A. A. *et al.* An aerial view of 80 years of climate-related glacier fluctuations in southeast Greenland. *Nature Geosci.* **5**, 427–432 (2012).
16. Cook, A., Fox, A. J., Vaughan, D. G. & Ferrigno, J. G. Retreating glacier-fronts on the Antarctic Peninsula over the last 50 years. *Science* **308**, 541–544 (2005).
17. Frezzotti, M., Cimbelli, A. & Ferrigno, J. G. Ice-front change and iceberg behaviour along the Oates and George V Coasts, Antarctica, 1912–1996. *Ann. Glaciol.* **27**, 643–650 (1998).
18. Frezzotti, M. & Polizzi, M. 50 years of ice-front changes between Adélie and Banzare Coasts, East Antarctica. *Ann. Glaciol.* **34**, 235–240 (2002).
19. Ferraccioli, F. *et al.* Aeromagnetic exploration over the East Antarctic ice sheet: a new view of the Wilkes subglacial basin. *Tectonophysics* **478**, 62–77 (2009).
20. van der Veen, C. J. Fracture mechanics approach to penetration of surface crevasses on glaciers. *Cold Reg. Sci. Technol.* **27**, 31–47 (1998).
21. Scambos, T. A., Hulbe, C., Fahnestock, M. J. & Bohlander, J. The link between climate warming and break-up of ice shelves in the Antarctic Peninsula. *J. Glaciol.* **46**, 516–530 (2000).
22. Hall, A. & Visbeck, M. Synchronous variability in the southern hemisphere atmosphere, sea ice, and ocean resulting from the southern annular mode. *J. Clim.* **15**, 3043–3057 (2002).
23. Thompson, D. W. J. & Solomon, S. Interpretation of recent Southern Hemisphere climate change. *Science* **296**, 895–899 (2002).
24. Massom, R. A. *et al.* Fast ice distribution in Adélie Land, East Antarctica: interannual variability and implications for emperor penguins *Aptenodytes forsteri*. *Mar. Ecol. Prog. Ser.* **374**, 243–257 (2009).
25. Jacobs, S. Observations of change in the Southern Ocean. *Phil. Trans. R. Soc. A* **364**, 1657–1681 (2006).
26. Arblaster, J. M. & Meehl, G. A. Contributions of external forcings to southern annular mode trends. *J. Clim.* **19**, 2896–2905 (2006).
27. de la Mare, W. K. Abrupt mid-twentieth century decline in Antarctic sea-ice extent from whaling records. *Nature* **389**, 57–60 (1997).
28. Parkinson, C. L. & Cavalieri, D. J. Antarctic sea ice variability and trends, 1979–2010. *Cryosphere* **6**, 871–880 (2012).
29. Amundson, J. M. *et al.* Ice mélange dynamics and implications for terminus stability, Jakobshavn Isbrae, Greenland. *J. Geophys. Res.* **115**, F01005 (2010).
30. Rignot, E., Mouginot, J. & Scheuchl, B. Ice flow of the Antarctic ice sheet. *Science* **333**, 1427–1430 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements Landsat imagery was provided free of charge by the US Geological Survey Earth Resources Observation Science Centre. We thank H. Pritchard for supplying data on glacier thinning.

Author Contributions C.R.S. and A.V. had the idea for the research. B.W.J.M. designed and undertook the mapping and data collection, and led the climate analysis. N.J.C. led the statistical analysis and all authors contributed to the analysis and interpretation of the results. C.R.S. wrote the first draft of the paper and all authors contributed to writing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.R.S. (c.r.stokes@durham.ac.uk).

METHODS

Data sources and glacier terminus mapping. We use optical satellite imagery to map the terminus position of 175 neighbouring outlet glaciers along the coast of East Antarctica. To remove intra- and interannual variability in glacier terminus positions on short timescales, we mapped glacier change at the end of the austral summer and at approximately decadal time steps spanning the past five decades. Our primary source of data was approximately 300 Landsat satellite images. Ideally, it would have been possible to pre-select the years of measurement, but our requirement for a comprehensive sample of glaciers spanning 5,400 km of coastline meant that only four main time steps allowed cloud-free mapping of the vast majority of their termini: 1974, 1990, 2000 and 2010. A small subsample of glaciers (38) were also mapped with declassified Argon imagery³¹ from 1963 (Supplementary Table 1), but few of these could be measured in 1974 and they were widely spread geographically. For these reasons (together with the lower resolution of the imagery in 1963), we exclude this epoch from more detailed analysis.

The absolute positional accuracy of the mapping was limited by co-registration of imagery from different sources and is measured at ± 75 to ± 210 m for Landsat imagery and up to ± 420 m for some Argon imagery. Overall, 85% of measurements have an error below ± 180 m, which is comparable to a similar study from the Antarctic Peninsula⁶, and is more than sufficient for extracting the decadal trends we present (Supplementary Table 1). To account for uneven changes along the calving front, glacier termini were digitized within a reference box that approximately delineated the sides of the glacier². The mean retreat distance was calculated as the area change at each time step, divided by the glacier width, which was obtained from the reference box.

Glacier flow speed and elevation change. We compare our data on glacier terminus change and width with recent measurements of their mean flow speed^{30,32} and elevation changes¹ (Supplementary Fig. 1). To guide the extraction of flow speeds and elevation changes, we used a map of grounding line positions using data from differential satellite synthetic-aperture radar interferometry³³. Ice velocity data were obtained from a high-resolution digital mosaic of ice motion in Antarctica^{30,32}, assembled from multiple satellite synthetic-aperture radar interferometry data acquired during the International Polar Year 2007 to 2009. This data set was used to extract a mean flow speed near the glacier terminus by digitizing an approximately square polygon that covered the width of the glacier and a similar distance in the flow direction (typically producing a box a few kilometres long and wide). The mean velocity was then extracted using the 'spatial analyst' tool in ARCGIS. Given the strong correlation between glacier velocity both up-ice and down-ice from the grounding line ($r^2 = 0.85$), and the fact that grounding line data were missing from some glaciers, all measurements were made close to the calving front. The glacier change data are suited to presentation on a cube-root scale in Figs 2 and 3, which allows informative display of long-tailed distributions, including both large positive and large negative values, and is consistent with the display of medians, quartiles, and 5th and 95th percentiles.

For comparison with recent measurements of elevation change¹ (2003–2007), we also mapped a sample of glaciers in austral summer 2006–2007 (see Supplementary Information and Supplementary Fig. 7). Data on ice elevation changes (thickening and thinning) were obtained from previously published ICESat laser altimetry¹

along the entire grounded margins of the Antarctic ice sheet between 2003 and 2007. Polygons were digitized immediately up-ice from the grounding line of those glaciers where data were available, and this permitted the extraction of elevation change data for 24% of glaciers (Supplementary Fig. 7a).

Climate data and statistical tests. Monthly mean surface air temperature records from four research stations in our study region (Scott, Dumont d'Urville, Casey and Mirny; Fig. 1) were extracted from the SCAR Met READER project (<http://www.antarctica.ac.uk/met/READER/>). All stations have complete monthly records that coincide with our glacier measurements between 1963 and 2010, apart from Scott, which has some data missing in 1994. Monthly values of the SAM index (Fig. 3) were obtained from <http://www.nerc-bas.ac.uk/icd/gima/sam.html>.

We determined whether the trends in glacier terminus position from the different epochs were statistically significant. The key issue is whether any observed differences between two epochs (for example 1974–1990 versus 1990–2000) are consistent with random variation at each epoch or whether they represent genuine differences between epochs. When data are normally distributed, this can be determined using a Student's *t*-test, which calculates the probability (*P* value) that differences as large as or larger than that observed could occur if the two sets being compared are not different. We follow the common conventions that $P < 0.05$ indicates a 'significant' difference, $P < 0.01$ indicates a 'highly significant' difference (99% confidence) and $P < 0.001$ indicates a 'very highly significant' difference.

Glacier data within each epoch are positively skewed towards a few very high values. Although the *t*-test is generally thought to be insensitive to violations of normality³⁴, especially with large sample sizes, and is unlikely to lead to a type-1 error (that is, find a significant difference that does not exist); it is prudent to use a non-parametric alternative: Wilcoxon's test, which does not assume normality. We performed two tests on data from each epoch: a 'paired' test, using only data from glaciers measured in both epochs (the Wilcoxon signed-rank test); and an 'unpaired' test, where data were included even if the glacier was measured only in one of the epochs (the Wilcoxon rank-sum test or Mann–Whitney *U* test). In unpaired *t*-tests, we follow standard procedure and allow unequal variances (heteroscedasticity). Results are presented in Supplementary Tables 2, 3, 5 and 6.

We also determined whether there were significant differences in mean austral summer temperatures (December–February) between the 1974–1990 and 1991–2000 epochs and the 1991–2000 and 2001–2010 epochs. In this case, normality holds to a good approximation, and so *t*-tests were performed on unpaired samples but were allowed unequal variances. Results are presented in Supplementary Table 4.

31. Kim, K., Jezek, K. C. & Liu, H. Orthorectified image mosaic of Antarctica from 1963 Argon satellite photography: image processing and glaciological applications. *Int. J. Remote Sens.* **28**, 5357–5373 (2007).
32. Rignot, E., Mouginot, J. & Scheuchl, B. MEaSUREs InSAR-based velocity map [version 1.0]. *National Snow and Ice Data Center* (2011).
33. Rignot, E., Mouginot, J. & Scheuchl, B. Antarctic grounding line mapping from differential satellite radar interferometry. *Geophys. Res. Lett.* **38**, L10504 (2011).
34. Blair, R. & Higgins, J. T. A comparison of the power of Wilcoxon's rank-sum statistic to that of the Student's *t* statistic under various non-normal distributions. *J. Educ. Stat.* **5**, 309–335 (1980).

Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage

Mohamed F. Haroon^{1*}, Shihu Hu^{2*}, Ying Shi², Michael Imelfort^{1,2}, Jurg Keller², Philip Hugenholtz^{1,3}, Zhiguo Yuan² & Gene W. Tyson^{1,2}

Anaerobic oxidation of methane (AOM) is critical for controlling the flux of methane from anoxic environments. AOM coupled to iron¹, manganese¹ and sulphate² reduction have been demonstrated in consortia containing anaerobic methanotrophic (ANME) archaea. More recently it has been shown that the bacterium *Candidatus 'Methyloirabilis oxyfera'* can couple AOM to nitrite reduction through an intra-aerobic methane oxidation pathway³. Bioreactors capable of AOM coupled to denitrification have resulted in the enrichment of '*M. oxyfera*' and a novel ANME lineage, ANME-2d^{4,5}. However, as '*M. oxyfera*' can independently couple AOM to denitrification, the role of ANME-2d in the process is unresolved. Here, a bioreactor fed with nitrate, ammonium and methane was dominated by a single ANME-2d population performing nitrate-driven AOM. Metagenomic, single-cell genomic and metatranscriptomic analyses combined with bioreactor performance and ¹³C- and ¹⁵N-labelling experiments show that ANME-2d is capable of independent AOM through reverse methanogenesis using nitrate as the terminal electron acceptor. Comparative analyses reveal that the genes for nitrate reduction were transferred laterally from a bacterial donor, suggesting selection for this novel process within ANME-2d. Nitrite produced by ANME-2d is reduced to dinitrogen gas through a syntrophic relationship with an anaerobic ammonium-oxidizing bacterium, effectively outcompeting '*M. oxyfera*' in the system. We propose the name *Candidatus 'Methanoperedens nitroreducens'* for the ANME-2d population and the family *Candidatus 'Methanoperedenaceae'* for the ANME-2d lineage. We predict that '*M. nitroreducens*' and other members of the '*Methanoperedenaceae*' have an important role in linking the global carbon and nitrogen cycles in anoxic environments.

Methanoperedenaceae fam. nov.

Candidatus 'Methanoperedens nitroreducens' gen. et sp. nov.

Etymology. *Methanoperedens*. *methano* (new Latin): pertaining to methane; *peredens* (Latin): consuming, devouring; *nitroreducens*. *nitro* (new Latin): pertaining to nitrate; *reducens* (Latin): leading back, bringing back and in chemistry converting to a different oxidation state. The name implies an organism capable of consuming methane and also reducing nitrogen-related compounds.

Locality. Enriched from a mixture of freshwater sediment and anaerobic wastewater sludge in Brisbane, Australia.

Diagnosis. Anaerobic, methane-oxidising, nitrate-reducing archaeon belonging to the *Methanoperedenaceae* fam. nov. (-aceae ending to denote a family). Grows as irregular cocci 1–3 µm in diameter and is typically found as sarcina-like clusters. Growth occurs at mesophilic temperatures (22–35 °C)⁵, pH 7–8 (refs 4, 5).

Microorganisms are central to the global methane cycle. Approximately 300 teragrams (Tg) of methane are produced annually through methanogenesis⁶, of which 90% is consumed by specialized groups of

archaea, collectively called ANME, that anaerobically oxidize methane before it diffuses into oxic environments⁷. AOM is proposed to occur by reversal of the canonical methanogenesis pathway based on incomplete metabolic reconstructions from three partial ANME-1 genomes^{8–10}, protein immunodetection¹¹ and *in vitro* enzymatic assays¹². However, validation of reverse methanogenesis in ANME remains to be demonstrated and has been hampered by an inability to obtain a pure culture.

AOM has been studied mostly in ANME-1 and ANME-2 that work in concert with sulphate-reducing bacteria (SRB); a process by which methane is oxidized to carbon dioxide and sulphate is reduced to hydrogen sulphide⁶. Recently, it has been suggested that ANME can couple AOM to sulphate reduction directly¹³. AOM can also be coupled to denitrification, as shown by the discovery of nitrite-driven aerobic methane oxidation in '*M. oxyfera*'³. ANME-2d has previously been co-enriched with '*M. oxyfera*' in bioreactors fed with methane and nitrate, raising the possibility that this ANME group may be involved in AOM coupled to nitrate reduction¹⁴. Given the prevalence of nitrate in the environment, verification of nitrate-driven AOM (as opposed to sulphate or nitrite as terminal electron acceptors) may have important implications for global methane cycling.

Here, an anaerobic laboratory-scale bioreactor was pulse-fed with nitrate, ammonium and methane for 350 days (see rationale below). Process data indicated simultaneous AOM, nitrate reduction and anammox (Fig. 1a). Fluorescence *in situ* hybridization (FISH) and 16S rRNA gene amplicon profiling revealed that the bioreactor community was dominated by ANME-2d (78% relative abundance; Fig. 1b, Supplementary Table 1 and Supplementary Fig. 1) making it a likely candidate responsible for the observed AOM. Notably, '*M. oxyfera*' and known SRB were not detected in the bioreactor community (Supplementary Table 1), eliminating the possibility that these populations contribute to nitrite- or sulphate-driven AOM, respectively. Organisms related (99% sequence identity) to the anammox bacterium, *Candidatus 'Kuenenia stuttgartiensis'*¹⁵, were present as a flanking population (approximately 3% relative abundance).

To confirm the predicted role of ANME-2d, we performed metagenomic sequencing of the bioreactor community. We assembled 13.3 gigabase pairs (Gb) of 150-bp paired-end Illumina HiSeq2000 data into 18,805 contigs (Supplementary Table 2) and separated the ANME-2d population from the flanking community based on guanine–cytosine content, read coverage and tetranucleotide frequencies (Supplementary Table 3 and Supplementary Figs 2–4). High relative abundance and a distinctive archaeal genome signature expedited this process. Six other genomes including the *Kuenenia* population could be resolved similarly (Supplementary Fig. 4). The ANME-2d contigs were reduced to 10 scaffolds using 3-kb-insert mate-pair sequence data (Supplementary Figs 5 and 6). Single-copy gene analysis confirmed completeness and fidelity of the genome (Supplementary Table 4). The genome size and guanine–cytosine content are 3.2 megabase pairs

¹Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Queensland 4072, Australia. ²Advanced Water Management Centre, Faculty of Engineering, Architecture and Information Technology, The University of Queensland, Brisbane, Queensland 4072, Australia. ³Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland 4072, Australia.

*These authors contributed equally to this work.

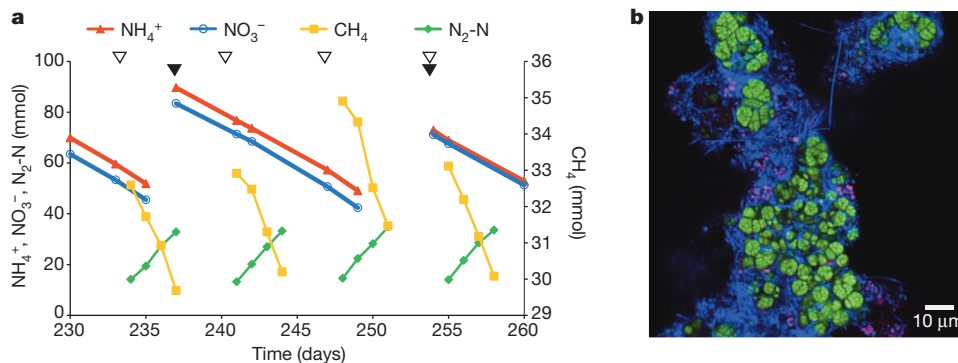


Figure 1 | Bioreactor performance data and microbial community composition. a, Key bioreactor performance data during steady-state operation from day 230 to 260 after inoculation. The arrows indicate pulse-feeding events of nitrate and ammonium (black), and methane (white). Nitrate, ammonium and methane consumption, and dinitrogen gas production, can be

(Mb) and 43%, respectively, comparable to a previously sequenced ANME-1b genome⁹ (Supplementary Table 5 and Supplementary Fig. 6). Based on the metabolic reconstruction of the obtained genome, we propose the name *Candidatus 'Methanoperedens nitroreducens'* and *Candidatus 'Methanoperedenaceae'* fam. nov. for the ANME-2d lineage (Supplementary Fig. 7).

A complete reverse methanogenesis pathway including all *mcr* subunit genes (*mcrABCDG*) and F420-dependent 5, 10-methenyltetrahydromethanopterin reductase (*mer*) genes were identified in the '*M. nitroreducens*' genome (Fig. 2). To our knowledge, this is the first report of a full reverse methanogenesis pathway in an ANME organism. The presence of a complete reductive acetyl-CoA (carbon fixation) pathway and acetyl-CoA synthetase (Fig. 2) suggest that '*M. nitroreducens*' may be able to produce acetate, as predicted for ANME-1 (refs 8, 9). We also identified genes for nitrate reduction but not subsequent steps in denitrification. The membrane-bound

observed. The concentration of nitrite was negligible throughout the experiment. **b**, Fluorescence *in situ* hybridization micrograph of the bioreactor community showing the dominant '*M. nitroreducens*' population in large, dense clusters (green), smaller flanking *Kuenenia* cells (magenta) and other bacteria (blue).

nitrate reductase is comprised of a molybdopterine oxidoreductase domain and twin-arginine signal peptide for translocation across the membrane (*narG*) and iron-sulphur centres (*narH*). The *narGH* genes seem to have been acquired laterally from a bacterial donor (Supplementary Figs 6, 8 and 9). To confirm that the inferred lateral transfer was not the result of a chimaeric archaeal–bacterial assembly, we validated the metagenomic scaffolds using mate-pair and paired-end read information (Supplementary Figs 5 and 10). Furthermore, we sequenced a single '*M. nitroreducens*' cell isolated and identified previously using 16S rRNA analysis¹⁶, and mapped individual and assembled single-cell reads to the near-complete '*M. nitroreducens*' population genome, confirming the fidelity of the original assembly, including the region of putative lateral gene transfer (Supplementary Table 6 and Supplementary Fig. 6). The '*M. nitroreducens*' *narGH* originated from the Proteobacteria (Supplementary Figs 8 and 9), adding further weight to previous reports that the *nar* complex can be transferred laterally between phylogenetic domains¹⁷. Codon usage profiles show that the introduced genes have adapted to their archaeal host (Supplementary Table 7 and Supplementary Fig. 11), suggesting that the identified lateral transfer occurred well before the establishment of the population in the bioreactor.

Metatranscriptomics and isotope labelling experiments were carried out on the bioreactor and subsamples of the biomass, respectively, to confirm the expected expression of metabolic pathways and fate of methane and nitrate in the system. All genes necessary for oxidation of methane to carbon dioxide were highly expressed relative to house-keeping genes in '*M. nitroreducens*' (Fig. 2 and Supplementary Table 4), consistent with process data (Fig. 1a). No other genes or transcripts for anaerobic or aerobic methane oxidation¹⁸ were detected in the metagenome or metatranscriptome. Further experiments confirmed that methane fed to the batch reactor was converted to carbon dioxide (Supplementary Fig. 12) and the added ¹³C-labelled methane was converted to ¹³C-labelled carbon dioxide (Fig. 3a). The MCR-specific inhibitor, 2-bromoethanesulphonate (BES), has been shown to inhibit AOM at low concentration (1 mM)¹⁹. However, BES inhibition tests using high concentrations (20 and 50 mM) did not have measurable effects on the AOM rates (Supplementary Fig. 13). The inability of BES to inhibit AOM in ANME is consistent with previous reports^{20,21} and is potentially explained by impermeability of the ANME cell wall as has previously been proposed for cultured methanogens²².

Concomitant with the high expression of genes responsible for reverse methanogenesis in the bioreactor, the *narGH* genes belonging to '*M. nitroreducens*' were also highly expressed relative to housekeeping genes. A small number of *narGH* homologues identified in the metagenome belonging to flanking populations were not highly expressed indicating that '*M. nitroreducens*' performed the great majority of nitrate reduction in the bioreactor (Fig. 1a, Supplementary Table

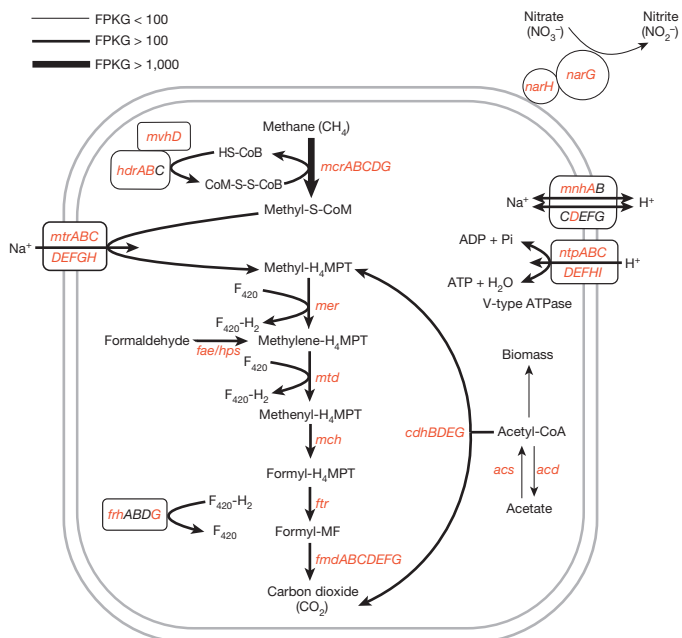


Figure 2 | Key carbon and nitrogen transformations in ‘*Methanoperedens nitroreducens*’. Reverse methanogenesis pathway in ‘*M. nitroreducens*’ coupled by an unknown electron carrier to nitrate reduction. Highly expressed genes are shown in red, indicating that the complete reverse methanogenesis pathway and nitrate reduction genes were active in the bioreactor. Increasing line thickness indicates increasing absolute gene-expression values. FPKG (fragments mapped per kilobase of gene length) is a measure of normalized gene expression.

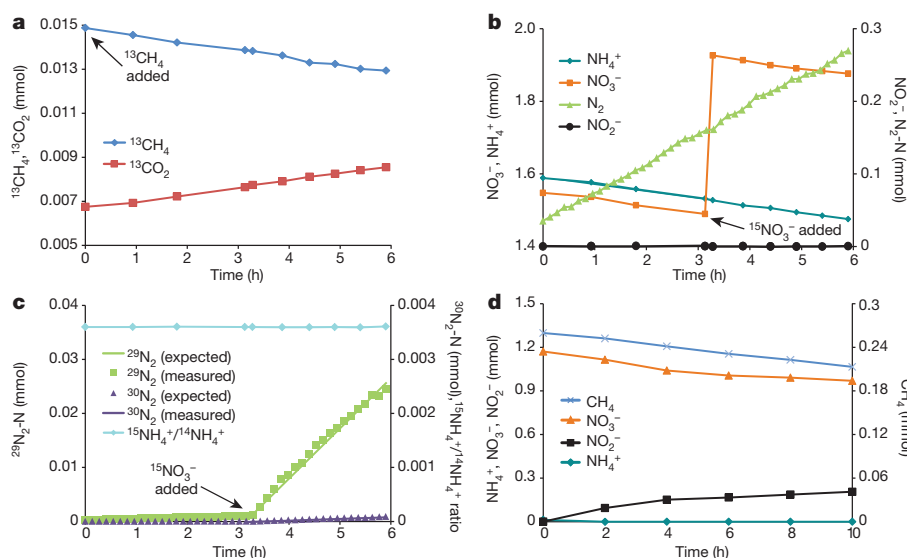
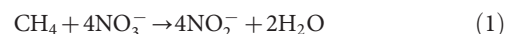


Figure 3 | Methane oxidation coupled to nitrate reduction in the AOM bioreactor. **a, b,** Data from isotope labelling batch test demonstrating stoichiometrically balanced conversion of $^{13}\text{CH}_4$ to $^{13}\text{CO}_2$ (**a**), and NO_3^- and NH_4^+ to N_2 without accumulation of NO_2^- (**b**). **c,** Experimentally measured production of $^{29}\text{N}_2$ and $^{30}\text{N}_2$ matching predictions (green and purple lines, respectively) based on assumed reactions (see details in Methods). The ratios between $^{15}\text{NH}_4^+$ and $^{14}\text{NH}_4^+$ are shown to confirm that there was no

8). To determine whether the *Kuenenia* anammox population uses the nitrite produced by '*M. nitroreducens*', a subsample of the bioreactor was fed unlabelled nitrate and ammonium (Fig. 3b) before addition of ^{15}N -labelled nitrate (Fig. 3c). In this scenario, we expected accumulation of ^{29}N -labelled dinitrogen gas as a result of anaerobic oxidation of $^{14}\text{NH}_4^+$ with $^{15}\text{NO}_2^-$. Mass spectrometry results confirmed the production of ^{29}N -labelled dinitrogen gas, according to reaction stoichiometry (Fig. 3c). The key genes for anaerobic ammonium oxidation by

conversion of $^{15}\text{NO}_3^-$ to $^{15}\text{NH}_4^+$. ^{30}N -labelled dinitrogen gas also increased as predicted, albeit to a much smaller extent, owing to reaction between $^{15}\text{NO}_2^-$ and naturally present $^{15}\text{NH}_4^+$. The ratios between $^{15}\text{NO}_3^-$ and $^{14}\text{NO}_3^-$ in the system before and after $^{15}\text{NO}_3^-$ addition were stable (data not shown). **d,** Batch test showing the accumulation of nitrite as a result of nitrate reduction when ammonium was removed from the feed.

the flanking *Kuenenia* population were highly expressed relative to their housekeeping genes (Supplementary Table 9). In further tests without ammonium feeding, we observed accumulation of nitrite in a subsample of the bioreactor (Fig. 3d), reconfirming the role of anammox in nitrite removal. Taken together, these data provide compelling evidence that '*M. nitroreducens*' couples AOM to nitrate reduction to nitrite, (equation (1)²³).



$$(\Delta G^\circ = -503 \text{ kJ mol}^{-1} \text{ CH}_4)$$

The original rationale for operating the bioreactor with ammonium was based on the hypothesis that anammox bacteria would out-compete '*Methylomirabilis oxyfera*' for nitrite according to known reaction rates²⁴, thus removing the main competitor to '*M. nitroreducens*' for methane and simplifying the analysis of AOM (Fig. 4a). Multiple lines of evidence (isotope labelling, electron and mass balance calculations, and metatranscriptomic data) confirm this hypothesis and indicate that in the presence of ammonium, nitrite produced by '*M. nitroreducens*' is used as an electron acceptor for ammonium oxidation to dinitrogen gas by the flanking *Kuenenia* population

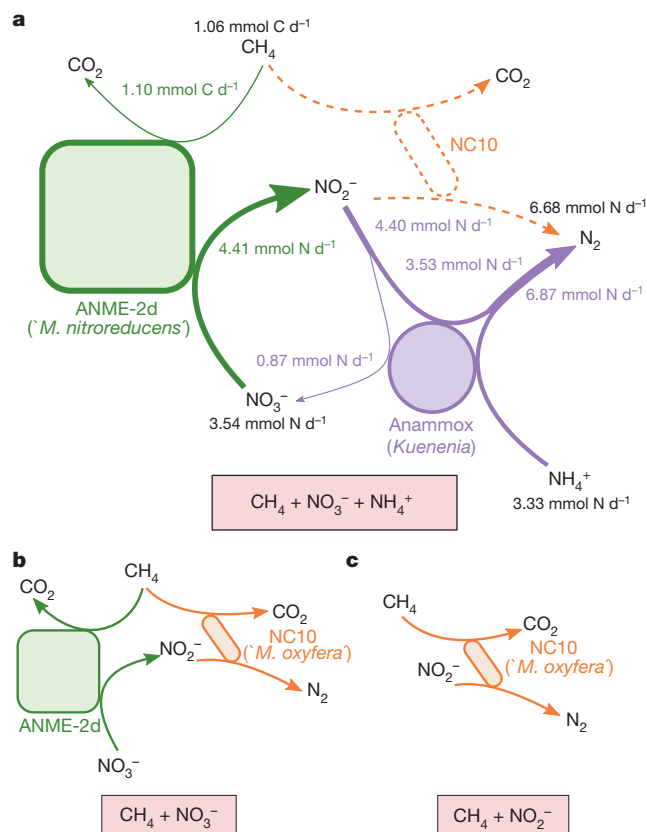


Figure 4 | Observed interactions between key populations in methane-fed bioreactors with differing nitrogen sources. Each population and its associated metabolic activities are colour-coded. **a,** Interactions under methane, nitrate and ammonium conditions in the current study. The thicknesses of arrows indicate reaction flux rates, and dashed arrows represent interactions not occurring under the given conditions (see Supplementary Table 10 for mass balance calculations). The methane consumption ($1.10 \text{ mmol C d}^{-1}$) and dinitrogen gas production ($6.87 \text{ mmol N d}^{-1}$) rates predicted based on measured ammonium, nitrate and nitrite data closely match their measured rates ($1.06 \text{ mmol C d}^{-1}$ and $6.68 \text{ mmol N d}^{-1}$) with only 3.8% and 2.8% errors, respectively. Mass and electron balance and molecular data further indicate that all of the methane consumed was used for reduction of nitrate to nitrite by '*M. nitroreducens*', and all of the produced nitrite ($4.41 \text{ mmol N d}^{-1}$) was consumed by *Kuenenia* ($4.40 \text{ mmol N d}^{-1}$). The outcompeted NC10 population is shown in dashed lines with no shading. **b,** Interactions in a bioreactor fed with methane and nitrate resulting in a co-culture of ANME-2d and '*M. oxyfera*'. **c,** Interactions in a bioreactor fed with methane and nitrite resulting in a dominant '*M. oxyfera*' population^{3,14}.

(Figs 1a and 4a, Supplementary Table 9 and Supplementary Table 10). This partnership is also beneficial for '*M. nitroreducens*' as *Kuenenia* partially replenishes the nitrate (Fig. 4a). Consistent with this partnership, we found spatial co-localization of '*M. nitroreducens*' and the *Kuenenia* population (Fig. 1b and Supplementary Fig. 14). In the absence of ammonium, '*M. oxyfera*' would not have to compete with *Kuenenia* leading to the previously reported co-culture of a ANME-2d species and '*M. oxyfera*'²⁴ (Fig. 4b). With nitrite as the only electron acceptor, '*M. nitroreducens*' is unable to compete with '*M. oxyfera*' (Fig. 4c), leading to a community dominated by the latter organism¹⁴. The significance of these different experimental scenarios in natural environments remains to be determined. However, our data conclusively demonstrate that '*M. nitroreducens*' is capable of coupling nitrate reduction to AOM as evidenced by closed electron and mass balance calculations and sole attribution of this metabolic activity to '*M. nitroreducens*' in the simplified bioreactor community.

METHODS SUMMARY

A 5.6 litre bioreactor fed with nitrate, ammonium and methane was inoculated with a microbial consortium performing nitrate-dependent AOM and another performing anammox. Bioreactor operation was carried out as described previously¹⁴. For batch tests, subsamples of the bioreactor biomass were transferred to 330-ml vessels, which were provided with methane, nitrate and in most cases ammonium. In isotopic batch tests, ¹³C-labelled methane and ¹⁵N-labelled nitrate were added. ¹³CH₄ and ¹³CO₂ were measured with gas chromatography-mass spectrometry and isotope ratio mass spectrometry. ¹⁵NO₃⁻ and ¹⁵NH₄⁺ were measured with a continuous flow elemental analyser-isotope ratio mass spectrometer²⁵. ²⁸N₂, ²⁹N₂ and ³⁰N₂ were measured with a membrane inlet mass spectrometer³.

Bioreactor microbial community composition was determined using 16S rRNA gene amplicon pyrosequencing²⁶ and fluorescence *in situ* hybridization²⁷. Metagenomic data was generated from a paired-end library sequenced on an Illumina HiSeq2000. The '*M. nitroreducens*' genome was binned using guanine-cytosine content, average contig coverage, and an emergent self-organizing map based on contig tetranucleotide frequencies²⁸. The binned contigs were scaffolded using a 3-kb-insert mate-pair library sequenced on an Ion Torrent Personal Genome Machine. Annotation of the genome was performed using the Integrated Microbial Genomes Expert Review system²⁹ and curated manually. To validate laterally transferred regions in the '*M. nitroreducens*' genome, single cells were sorted, whole-genome amplified¹⁶ and sequenced on Illumina GAII. Individual and assembled reads from the single cell were mapped against the '*M. nitroreducens*' genome and coverage values were determined. For metatranscriptomics, total RNA was extracted, ribosomal RNA subtracted, and complementary DNA synthesized (as described previously³⁰ with modifications) before sequencing on an Illumina MiSeq. Reads were mapped against the metagenome and gene expression was determined by calculating the number of fragments mapped per kilobase of gene length (FPKG). Highly expressed genes were identified by comparing FPKG values to those of housekeeping genes (Supplementary Table 4).

Full Methods and any associated references are available in the online version of the paper.

Received 2 October 2012; accepted 11 June 2013.

Published online 28 July 2013.

1. Beal, E. J., House, C. H. & Orphan, V. J. Manganese- and Iron-Dependent Marine Methane Oxidation. *Science* **325**, 184–187 (2009).
2. Boetius, A. *et al.* A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature* **407**, 623–626 (2000).
3. Ettwig, K. F. *et al.* Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature* **464**, 543–548 (2010).
4. Raghoebarsing, A. A. *et al.* A microbial consortium couples anaerobic methane oxidation to denitrification. *Nature* **440**, 918–921 (2006).
5. Hu, S. *et al.* Enrichment of denitrifying anaerobic methane oxidizing microorganisms. *Environmental Microbiol. Rep.* **1**, 377–384 (2009).
6. Knittel, K. & Boetius, A. Anaerobic oxidation of methane: progress with an unknown process. *Annu. Rev. Microbiol.* **63**, 311–334 (2009).
7. Reeburgh, W. S. Oceanic Methane Biogeochemistry. *Chem. Rev.* **107**, 486–513 (2007).
8. Hallam, S. J., Girguis, P. R., Preston, C. M., Richardson, P. M. & DeLong, E. F. Identification of methyl coenzyme M reductase A (*mcrA*) genes associated with methane-oxidizing archaea. *Appl. Environ. Microbiol.* **69**, 5483–5491 (2003).
9. Meyerdieters, A. *et al.* Metagenome and mRNA expression analyses of anaerobic methanotrophic archaea of the ANME-1 group. *Environ. Microbiol.* **12**, 422–439 (2010).

10. Stokke, R., Roalkvam, I., Lanzen, A., Hafflidason, H. & Steen, I. H. Integrated metagenomic and metaproteomic analyses of an ANME-1-dominated community in marine cold seep sediments. *Environ. Microbiol.* **14**, 1333–1346 (2012).
11. Heller, C., Hoppert, M. & Reiter, J. Immunological localization of coenzyme M reductase in anaerobic methane-oxidizing archaea of ANME 1 and ANME 2 type. *Geomicrobiol. J.* **25**, 149–156 (2008).
12. Scheller, S., Goenrich, M., Boecher, R., Thauer, R. K. & Jaun, B. The key nickel enzyme of methanogenesis catalyses the anaerobic oxidation of methane. *Nature* **465**, 606–608 (2010).
13. Milucka, J. *et al.* Zero-valent sulphur is a key intermediate in marine methane oxidation. *Nature* **491**, 541–546 (2012).
14. Hu, S., Zeng, R. J., Keller, J., Lant, P. A. & Yuan, Z. Effect of nitrate and nitrite on the selection of microorganisms in the denitrifying anaerobic methane oxidation process. *Environmental Microbiol. Rep.* **3**, 315–319 (2011).
15. Strous, M. *et al.* Deciphering the evolution and metabolism of an anammox bacterium from a community genome. *Nature* **440**, 790–794 (2006).
16. Yilmaz, S., Haroon, M. F., Rabkin, B. A., Tyson, G. W. & Hugenholtz, P. Fixation-free fluorescence *in situ* hybridization for targeted enrichment of microbial populations. *ISME J.* **4**, 1352–1356 (2010).
17. Cabello, P., Roldán, M. D. & Moreno-Vivián, C. Nitrate reduction and the nitrogen cycle in archaea. *Microbiology* **150**, 3527–3546 (2004).
18. Trotsenko, Y. A. & Murrell, J. C. in *Advances in Applied Microbiology* Vol. 63 (eds Sariaslani, S., Laskin, A. I. & Geoffrey, M. G.) 183–229 (Academic Press, 2008).
19. Nauhaus, K., Treude, T., Boetius, A. & Krüger, M. Environmental regulation of the anaerobic oxidation of methane: a comparison of ANME-I and ANME-II communities. *Environ. Microbiol.* **7**, 98–106 (2005).
20. Orcutt, B., Samarkin, V., Boetius, A. & Joye, S. On the relationship between methane production and oxidation by anaerobic methanotrophic communities from cold seeps of the Gulf of Mexico. *Environ. Microbiol.* **10**, 1108–1117 (2008).
21. Dekas, A. E., Poretsky, R. S. & Orphan, V. J. Deep-sea archaea fix and share nitrogen in methane-consuming microbial consortia. *Science* **326**, 422–426 (2009).
22. Smith, M. R. Reversal of 2-bromoethanesulfonate inhibition of methanogenesis in *Methanosarcina* sp. *J. Bacteriol.* **156**, 516–523 (1983).
23. Caldwell, S. L. *et al.* Anaerobic oxidation of methane: mechanisms, bioenergetics, and the ecology of associated microorganisms. *Environ. Sci. Technol.* **42**, 6791–6799 (2008).
24. Luesken, F. A. *et al.* Simultaneous nitrite-dependent anaerobic methane and ammonium oxidation processes. *Appl. Environ. Microbiol.* **77**, 6802–6807 (2011).
25. Brooks, P. D., Stark, J. M., McInteer, B. B. & Preston, T. Diffusion method to prepare soil extracts for automated nitrogen-15 analysis. *Soil Sci. Soc. Am. J.* **53**, 1707–1711 (1989).
26. Kunin, V., Engelbrektson, A., Ochman, H. & Hugenholtz, P. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ. Microbiol.* **12**, 118–123 (2010).
27. Glöckner, F. O. *et al.* An *in situ* hybridization protocol for detection and identification of planktonic bacteria. *System. Applied Microbiol.* **19**, 403–406 (1996).
28. Dick, G. J. *et al.* Community-wide analysis of microbial genome sequence signatures. *Genome Biol.* **10**, R85 (2009).
29. Markowitz, V. M. *et al.* IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* **25**, 2271–2278 (2009).
30. Stewart, F. J., Ottesen, E. A. & DeLong, E. F. Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *ISME J.* **4**, 896–907 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the ACE sequencing team; M. Butler, F. May and S. Low for their help with the 454 pyrosequencing, Illumina and Ion Torrent sequencing and the DOE Joint Genome Institute for single-cell sequencing. We also thank P. Lu for assistance with bioreactor operation, R. Zeng and P. Lant for their contribution to the development of initial bioreactors, B. Keller, J. Li, Y. Rui and Y. Wang for chemical and isotopic analyses, and D. Willner for assistance with genomic analysis. We are grateful to J. Euzéby for etymological advice. This work is supported by the Australian Research Council (ARC) through projects DP0666762 and DP0987204 and strategic funds from the Australian Centre for Ecogenomics. G.W.T. is supported by an ARC Queen Elizabeth II Fellowship (DP1093175). P.H. is supported by an ARC Discovery Outstanding Researcher Award (DP120103498). Y.S. is supported by the China Scholarship Council.

Author Contributions S.H. and Y.S. enriched the microorganisms in the parent bioreactor and performed batch and isotope labelling tests. S.H., Y.S., J.K. and Z.Y. performed the process data analysis. M.F.H. performed the sampling, preservation and nucleic acid extractions. M.F.H. prepared samples for single-cell genomics, metagenomic and metatranscriptomic sequencing. M.F.H. and G.W.T. performed the microbial community analysis. M.I. developed the bioinformatic tools. M.F.H., M.I., P.H. and G.W.T. performed the bioinformatic analyses. M.F.H., S.H., Z.Y., P.H., G.W.T. wrote the manuscript in consultation with all other authors.

Author Information Sequencing data are deposited at the NCBI Short Read Archive under accession numbers SRR925398, SRR925402 and SRR901892. Annotated assembled sequences were incorporated into the IMG system with the Taxon Object ID 2515154041. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.W.T. (g.tyson@uq.edu.au) or Z.Y. (z.yuan@awmc.uq.edu.au).

METHODS

Bioreactor operation, nutrient and gas measurements. The bioreactor was inoculated from two previously established bioreactors, one performing nitrate-dependent AOM⁵ and the other performing anammox with a nitrite, ammonium and methane feed. Using group-specific FISH probes (see FISH section below), the AOM inoculum was dominated by ANME-2d (60%) and *M. oxyfera* (30%), and the anammox bioreactor was dominated by *Kuenenia*-like planctomycetes (50%) and *M. oxyfera* (20%). At the time of sampling, the nitrate consumption rate of the AOM bioreactor was approximately $1.1 \text{ mmol l}^{-1} \text{ d}^{-1}$, the volatile suspended solids (VSS) concentration was approximately 0.8 g l^{-1} , the nitrite loading rate for the anammox bioreactor was approximately $1.4 \text{ mmol l}^{-1} \text{ d}^{-1}$, and the VSS concentration was approximately 1.1 g l^{-1} . Equal volumes of the two inocula (500 ml) were mixed with 3.6 l of medium⁴ in a 5.6 l glass bioreactor.

The bioreactor was operated at room temperature ($22 \pm 2^\circ \text{C}$) and mixed continuously with a magnetic stirrer at 200 r.p.m. To provide methane as the carbon and energy source, a mixed gas (5% N_2 , 90% CH_4 and 5% CO_2) was used to flush the headspace regularly to maintain the methane partial pressure between 0.5 and 1.0 atm. The pH of the bioreactor was monitored with a pH probe (TPS) and controlled between 7.0 and 7.5 by manual injection of a 1M HCl solution. Other operational conditions are as described previously¹⁴. The concentration of nitrate and ammonium in the bioreactor was maintained between 5.0 and 10.0 mmol l^{-1} by manual injection of concentrated stock solutions ($80 \text{ g NO}_3^- \text{--} \text{N l}^{-1}$ and $48 \text{ g NH}_4^+ \text{--} \text{N l}^{-1}$, respectively). Stock solutions were prepared with degassed milli-Q water and stored in sealed bottles with nitrogen in the headspace.

Fluorescence in situ hybridization. Cells from the bioreactor were collected and fixed in 4% paraformaldehyde before FISH²⁷. Bioreactor microbial communities were hybridized with the following oligonucleotide probes; ANME-2-specific probe (Darch872-FITC)⁴, anammox-specific probe (AMX820-Cy3)³¹, *M. oxyfera*-specific probe (DBACT193-Cy3)⁴ and a general bacterial probe (EUB338+-Cy5)³² (final concentration of 5 ng for each probe) for 2 h at 46°C . Labelled cells were visualized on an LSM512 confocal laser scanning microscope (Carl Zeiss) with Ar-ion laser (488 nm) and two HeNe lasers (543 and 633 nm). To determine the spatial arrangement of *M. nitroreducens* and anammox cells, 15 images were acquired at random locations in each well. All images were processed using the DAIME software³³ and split into individual colour channels before image segmentation. Cells were identified using the automatic segmentation in the DAIME and artefacts (irregular shapes and bright autofluorescence) were rejected manually. Segmentation was performed for all images using the 'batch processing' option. Spatial co-localization analysis was performed using the Linear Dipole algorithm in DAIME.

DNA extraction. DNA was extracted from 2 ml of bioreactor biomass using a FastDNA SPIN Kit for Soil (MP Biomedicals) with Fastprep beadbeater (Bio101) according to manufacturer's instructions. DNA concentrations were quantified using the Quant-iT dsDNA HS assay kit (Invitrogen) as per the manufacturer's instructions.

Microbial community profiling. 16S rRNA genes were amplified using the 'universal' primer set 926F/1392R²⁶. PCR products were purified using a QIAquick PCR Purification Kit (Qiagen) and quantified using Quant-iT dsDNA HS assay kit (Invitrogen). Amplicons were pooled in equimolar concentrations and sequenced using a 454 GS FLX Titanium sequencer (Roche) as per the manufacturer's protocol.

Amplicon sequences were processed using the ACE Pyrotag Pipeline (APP) v.2.3.2 (<http://github.com/ECogenomics/APP>). In brief, sequences were quality filtered and de-multiplexed with the split_libraries.py script of QIIME v.1.3.0 using default parameters³⁴ and then checked for chimaeras against Greengenes database³⁵ using UCHIME³⁶. Sequences less than 350 bp in length were removed. Homopolymeric errors were corrected using Acacia v.1.48 (ref. 37). Sequences with 97% identity were clustered as operational taxonomic units (OTUs) using UCLUST³⁸ and then taxonomically assigned using the Greengenes database.

Metagenomic sequencing and assembly. Community genomic DNA from the bioreactor was sheared to approximately 500 bp using a Covaris-S2 instrument (Covaris) and a paired-end library was prepared using the TruSeq PE Cluster Kit v3-cBot-HS and TruSeq SBS kit v3-HS sequencing kit (Illumina). The library was sequenced on the HiSeq 2000 (Illumina) platform generating 2×150 -bp paired-end reads with an average insert length of 250 bp.

Reads were filtered based on quality score and reads less than 100 bp in length were removed using CLC Genomic Workbench v.5.1.5 (CLC bio). Filtered sequences were assembled using CLC's *de novo* assembler using a *k*-mer length of 63. The average coverage of each contig > 1 kb was quantified by mapping all the reads using CLC Genomics Workbench with a minimum 95% similarity and 95% of the read length.

To separate the *M. nitroreducens* genome from other organisms in the metagenome, we used a combination of read coverage, guanine-cytosine content

and tetranucleotide frequencies²⁸ calculated using *kmer_counter* (http://github.com/wwood/bioruby-kmer_counter). Redundancy analysis was performed on these data and visualized as a principal component analysis plot using the R packages *vegan* and *ggplot2* respectively (<http://www.r-project.org/>). The metagenome was also binned using an emergent self-organizing map (ESOM; <http://www.cbs.dtu.dk/courses/27618.chemo/ESOM.pdf>) as described previously²⁸. In brief, tetranucleotide frequencies were clustered on a toroidal map with Euclidean grid distance and dimensions scaled to 160×268 . Training used the K-Batch algorithm ($k = 0.15\%$) for 20 training epochs. The standard best match search method was used with local best match radius of 8. Contig fragments with similar tetranucleotide patterns were clustered and different genome bins were manually identified as areas separated by distinct ridge lines.

To scaffold the binned *M. nitroreducens* contigs, a 3-kb-insert mate-paired library was prepared according to the Ion Mate-Paired Library protocol (Life Technologies) and subsequently sequenced using a Ion 316 chip on an Ion Torrent Personal Genome Machine (PGM) (Life Technologies). Forward and reverse mate-pairs were extracted and mapped against the binned contigs. Scaffolds were constructed based on the mate-pair links between the contigs using *ionPairer* (<http://github.com/ECogenomics/ionPairer>). To check the completeness and fidelity of the assembly, an essential single-copy gene analysis was performed using AMPHORA2 (ref. 39). The near-complete *M. nitroreducens* genome was submitted to IMG/ER²⁹ for gene calling and functional annotation. Coding sequences were assigned a unique identifier prefixed with 'ANME2D'. Annotations for the methane and nitrogen genes were confirmed manually. Codon usage profiles were calculated using CodonW (<http://codonw.sourceforge.net/>).

Single-cell genomic sequencing and assembly validation. A single amplified *M. nitroreducens* cell generated as part of a previous study¹⁶ was sequenced (75-bp PE reads) using an Illumina Genome Analyser II. Reads were mapped onto the *M. nitroreducens* genome and the bioreactor metagenomic contigs using *bwa* v.0.6.2-r126 (ref. 40) with default settings. Samtools v0.0.18 (ref. 41) was used to calculate the percentage of mapped reads. The *M. nitroreducens* scaffolds were cut into a total of 3,205 non-overlapping 1-kb windows and the log median coverage was calculated for each window using the *M. nitroreducens*-only mapping and a custom Perl script.

Phylogenetic inference. For protein-coding genes, homologues were identified in reference genomes available through IMG, and imported into ARB⁴². Amino acid sequences were aligned using ClustalW, and phylogenies were calculated using parsimony and maximum likelihood (Phylip, PROTPARS). For 16S rRNA gene phylogenetic analysis, sequences were aligned using NAST⁴³, and imported into ARB. Phylogenetic trees were constructed by using neighbour joining and maximum-likelihood methods.

For the genome tree, all finished and draft bacterial and archaeal genomes were downloaded from the IMG4.0 (<http://img.jgi.doe.gov/>). Hidden Markov models (HMMs) for a set of 38 essential single-copy genes were obtained from the authors of Phyllosift (<http://phyllosift.wordpress.com/>; Supplementary Table 4). Candidate single-copy conserved genes from the IMG genomes were identified and aligned to the HMMs using HMMER3 (ref. 44). Unaligned characters were masked out and the resulting sequences were concatenated. A phylogenetic tree with estimated branch support values was constructed from these concatenated alignments using FastTree v.2.1.3 (ref. 45) with default settings and visualized in ARB.

Metatranscriptomics. A subsample of the bioreactor biomass (approximately 10 ml) was collected and RNA was extracted using RNA PowerSoil Total RNA Isolation kit (MO-BIO) according to manufacturer's protocol. Turbo DNA-free kit (Ambion) was used to remove genomic DNA and purified using RNeasy MinElute cleanup kit (Qiagen). rRNA subtraction was performed using Ribo-Zero Magnetic kit (Epicentre) as per the manufacturer's instructions. rRNA subtracted total RNA was amplified using MessageAmp II Bacteria kit (Ambion) as described previously⁴⁶. Amplified RNA was reverse transcribed to cDNA using the SuperScript III First-Strand Synthesis System (Invitrogen) via random hexamer primers and SuperScript Double-Stranded cDNA synthesis kit (Invitrogen) as described previously³⁰ before sequencing library preparation. Libraries were prepared using an Illumina Nextera DNA sample preparation kit according to the manufacturer's instructions, with the following modifications. During the PCR cleanup with Agencourt AMPure XP magnetic beads (Beckman Coulter Genomics), the ratio of beads to sample was increased to 1.8:1 to capture smaller fragments. The library was then sequenced on the Illumina MiSeq (150 base pair paired end reads).

Sequences were mapped against the metagenome using BWA v.0.6.2 (ref. 40). For quantification of gene expression, the number of reads that mapped against a particular gene was normalized by the length of the gene to generate a number of fragments mapped per kilobase of gene length (FPKG) using the *sam2fpgk.pl* script v.0.1 (<http://github.com/minilinin/sam2fpgk>), which was subsequently

used to determine their expression relative to housekeeping genes (Supplementary Table 4).

Batch experiments. Isotopic labelling experiments were conducted in triplicate in 330-ml stirred glass vessels without headspace, enabling methane consumption to be easily determined. A subsample of bioreactor biomass (380 ml) was transferred to each vessel, with the excess 50 ml of biomass stored in a reservoir bottle connected to the lid of the vessel through a tube. The biomass was flushed with a mixed gas (90% CH₄, 5% CO₂, 5% N₂) for 30 min. Sixty millilitres of ¹³C-labelled methane gas (Sigma Aldrich, 99 atom % ¹³C) was injected through the septum, displacing the equivalent amount of biomass into the reservoir bottle. After 30 min, the gaseous ¹³C-labelled methane was removed from the system by displacement with the biomass from the reservoir bottle. Nitrate and ammonium were added to the vessel by injecting concentrated anaerobic stock solutions to reach a concentration between 3.6–7.1 mmol N l⁻¹. Three hours after the experiments started, ¹⁵N-labelled nitrate was injected to achieve a concentration of 1.4 mmol ¹⁵NO₃⁻ l⁻¹, which was approximately 20–30% of the total nitrate present. The experiment was then continued for 3 h. The headspace of the reservoir bottle was flushed with helium continuously during the experiment to keep it oxygen-free. Liquid samples were collected from the sampling port at the middle of the vessel and filtered using 0.22-µm filters and stored in vacuum tubes (BD Vacutainer). To compensate for liquid losses due to sampling, the valve between the vessel and the reservoir was opened during sampling, and the dilution effect of the extra biomass/liquid brought into the vessel was taken into account during data processing.

Isotopically labelled samples containing nitrate and ammonium were processed with a method described previously²⁵ and the isotopic fractions were measured by EA-IRMS with a Eurovector Elemental Analyzer (Isoprime-EuroEA 3000). The dissolved methane and carbon dioxide concentrations and their isotopic fractions were measured with GC-MS-IRMS. ²⁸N₂, ²⁹N₂ and ³⁰N₂ were measured in real-time using membrane inlet mass spectrometer (MIMS)³, the sum of which gave the total N₂ concentration. The measured ²⁸N₂, ²⁹N₂ and ³⁰N₂ were compared with their predicted values based on the assumption that nitrate was reduced to nitrite by '*M. nitroreducens*' and all the produced nitrite was removed by *Kuenenia* through the anammox reaction. The production rates of ²⁹N₂ and ³⁰N₂ dinitrogen gas after ¹⁵N-labelled nitrate addition were predicted as:

$$r^{29}\text{N}_2\text{-N} = {}^{15}\text{N}\% \text{ of NO}_3^- \times {}^{14}\text{N}\% \text{ of NH}_4^+ \times r\text{NH}_4^+ + {}^{14}\text{N}\% \text{ of NO}_3^- \times {}^{15}\text{N}\% \text{ of NH}_4^+ \times r\text{NH}_4^+$$

$$r^{30}\text{N}_2\text{-N} = {}^{15}\text{N}\% \text{ of NO}_3^- \times {}^{15}\text{N}\% \text{ of NH}_4^+ \times r\text{NH}_4^+$$

Batch experiments were also conducted in triplicate to observe the nitrogen conversion in the absence of ammonium. For each test, a subsample of bioreactor biomass was transferred to a stirred glass vessel (330 ml) after ammonium in the bioreactor was depleted. The biomass was flushed with the mixed gas (described above) for 30 min to saturate with methane. Nitrate was then added to reach a concentration of approximately 3.6 mmol l⁻¹. Each test was run for approximately 10 h. Liquid samples were taken every 2 h for the analysis of dissolved methane, ammonium, nitrite and nitrate concentrations as described previously¹⁴.

For BES inhibition tests, 260 ml of bioreactor biomass was added per 330 ml batch vessel, with methane, ammonium and nitrate added as described above. BES was added to two sets of triplicate vessels 24 h after initiation of the experiment to a final concentration of 20 mM and 50 mM, respectively. Methane, dinitrogen gas, ammonium, nitrite and nitrate concentrations were monitored with the method described previously¹⁴.

Mass and electron balance calculations. The consumption rate of methane (rCH₄), ammonium (rNH₄⁺), nitrate (rNO₃⁻) and nitrite (rNO₂⁻), and production rate of dinitrogen gas (rN₂-N) were determined from the respective measured

concentration profiles through linear regression (see Supplementary Table 10). Nitrogen balance was calculated as the difference between the measured dinitrogen gas production rate and the sum of the measured ammonium and nitrate consumption rates: rN₂-N - (rNH₄⁺ + rNO₃⁻). The percentage error was calculated as the ratio between nitrogen balance error and the measured dinitrogen gas production rate: (rN₂-N - (rNH₄⁺ + rNO₃⁻))/rN₂-N. Electron balance was calculated as the difference between the electrons required by the reduction of nitrate to dinitrogen gas and the electrons produced by the oxidation of ammonium to dinitrogen gas and methane to CO₂: -5 × rNO₃⁻ + 3 × rNH₄⁺ + 8 × rCH₄. The percentage error was calculated as the ratio between electron balance error and electrons required by the reduction of nitrate to dinitrogen gas: (-5 × rNO₃⁻ + 3 × rNH₄⁺ + 8 × rCH₄)/-5 × rNO₃⁻. The nitrite consumption rate and nitrate production rate of anammox were calculated based on the ammonium consumption rate (rNH₄⁺) and anammox reaction stoichiometry⁴⁷:



The net nitrate reduction rate was calculated as the sum of measured nitrate consumption rate (rNO₃⁻) and nitrate production rate of anammox reaction. The methane consumption rate was predicted based on the net nitrate reduction rate and equation (1), and compared with measured methane consumption rate (rCH₄). The dinitrogen gas production rate was predicted as the sum of the measured ammonium and nitrate consumption rates (rNH₄⁺ + rNO₃⁻), and compared with the measured dinitrogen gas production rate (rN₂-N).

- Schmid, M., Schmitz-Esser, S., Jetten, M. & Wagner, M. 16S-23S rDNA intergenic spacer and 23S rDNA of anaerobic ammonium-oxidizing bacteria: implications for phylogeny and in situ detection. *Environ. Microbiol.* **3**, 450–459 (2001).
- Daims, H., Brühl, A., Amann, R., Schleifer, K.-H. & Wagner, M. The domain-specific probe EUB338 is insufficient for the detection of all bacteria: development and evaluation of a more comprehensive probe set. *Systematic Applied Microbiol.* **22**, 434–444 (1999).
- Daims, H., Lückner, S. & Wagner, M. Daime, a novel image analysis program for microbial ecology and biofilm research. *Environ. Microbiol.* **8**, 200–213 (2006).
- Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nature Meth.* **7**, 335–336 (2010).
- McDonald, D. et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* **6**, 610–618 (2012).
- Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200 (2011).
- Bragg, L., Stone, G., Imelfort, M., Hugenholtz, P. & Tyson, G. W. Fast, accurate error-correction of amplicon pyrosequences using Acacia. *Nature Meth.* **9**, 425–426 (2012).
- Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
- Wu, M. & Scott, A. J. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. *Bioinformatics* **28**, 1033–1034 (2012).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
- Ludwig, W. et al. ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**, 1363–1371 (2004).
- DeSantis, T. Z. et al. NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Res.* **34**, W394–W399 (2006).
- Eddy, S. R. Accelerated profile HMM searches. *PLOS Comput. Biol.* **7**, e1002195 (2011).
- Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
- Shi, Y., Tyson, G. W. & DeLong, E. F. Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature* **459**, 266–269 (2009).
- Kuenen, J. G. Anammox bacteria: from discovery to application. *Nature Rev. Microbiol.* **6**, 320–326 (2008).

Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations

Gregory I. Lang^{1*†}, Daniel P. Rice^{2*}, Mark J. Hickman³, Erica Sodergren⁴, George M. Weinstock⁴, David Botstein¹ & Michael M. Desai²

The dynamics of adaptation determine which mutations fix in a population, and hence how reproducible evolution will be. This is central to understanding the spectra of mutations recovered in the evolution of antibiotic resistance¹, the response of pathogens to immune selection^{2,3}, and the dynamics of cancer progression^{4,5}. In laboratory evolution experiments, demonstrably beneficial mutations are found repeatedly^{6–8}, but are often accompanied by other mutations with no obvious benefit. Here we use whole-genome whole-population sequencing to examine the dynamics of genome sequence evolution at high temporal resolution in 40 replicate *Saccharomyces cerevisiae* populations growing in rich medium for 1,000 generations. We find pervasive genetic hitchhiking: multiple mutations arise and move synchronously through the population as mutational ‘cohorts’. Multiple clonal cohorts are often present simultaneously, competing with each other in the same population. Our results show that patterns of sequence evolution are driven by a balance between these chance effects of hitchhiking and interference, which increase stochastic variation in evolutionary outcomes, and the deterministic action of selection on individual mutations, which favours parallel evolutionary solutions in replicate populations.

Evolutionary adaptation is driven by the accumulation of beneficial mutations. The traditional view is that these dynamics are dominated by rare beneficial ‘driver’ mutations that occasionally survive drift and increase in frequency until they fix (a ‘selective sweep’)^{9,10}. This implicitly assumes that at most a single beneficial mutation is present in the population at once. However, recent experiments have shown that even for modestly sized populations of microbes and viruses, beneficial mutation rates are large enough^{11,12} that multiple driver mutations spread simultaneously, an effect known as ‘clonal interference’. This means that the fate of each mutation depends not only on its own effect on fitness but also on the rest of the variation in the population: neutral or deleterious mutations can fix if they occur in very fit genetic backgrounds, and beneficial mutations occurring in unfit lineages cannot succeed^{13–17}.

Recent work has uncovered important consequences of these clonal interference effects. For example, interference alters the rate of adaptation^{18,19}, the fate of marked lineages^{20,21}, and the distribution of fitness effects of fixed mutations^{12,16,17}. However, the underlying basis of these effects at the genomic sequence level has not been observed directly. A number of questions arise regarding the fate of those mutations that occur, the changes in frequency of each mutation over time, and the way in which these sequence-level dynamics determine the rate and repeatability of adaptation. Recent studies have sequenced clones or whole-population samples from microbial evolution experiments^{6,22–25}, but apart from studies in viral systems^{26–28}, this work has been limited to individual clones or populations or to widely separated time points that lack the temporal resolution to address these questions.

Here we describe the first direct and detailed view of the dynamics of genomic sequence evolution across many replicate microbial populations. In previous work²¹, we adapted approximately 600 replicate haploid yeast populations to growth in rich medium for 1,000 generations, half at ‘large’ (10^6) and half at ‘small’ (10^5) population sizes. Here we report the sequencing of whole-population samples from 40 of these populations (14 large and 26 small), chosen because we previously followed a single marker above a frequency of 0.1 (ref. 21). Each population was sequenced to 100-fold depth at 12 time points (approximately every 80 generations) for a total of 480 sequenced time points. Distinguishing mutations from sequencing errors in this whole-population sequence data is challenging. However, the high temporal resolution of our data permits the identification of mutations even at relatively low frequency by leveraging multiple time points. We developed two independent pipelines for this purpose, which rely on the fact that real mutations (but not sequencing or alignment errors) have frequencies that are correlated through time (Methods). This strategy allowed us to identify mutations that rose to a frequency of at least approximately 0.1 and to track these mutations through the rest of the timecourse. Across the 40 populations, we identified a total of 1,020 mutations, 253 of which fix; we annotated each to a gene or intergenic region and classified coding mutations as synonymous or nonsynonymous (Supplementary Table 1). Figure 1 shows six representative populations; the remaining populations exhibit similar patterns (Supplementary Fig. 1).

Averaged across all 40 populations, the rate at which mutations appeared and subsequently went extinct or fixed was constant through 1,000 generations (Fig. 2a). The average within-population polymorphism increased steadily through the first 600 generations, before saturating (Fig. 2a). In individual populations, however, the appearance of mutations is highly punctuated. This leads to the most striking feature of our results: selective sweeps are rarely single mutation or single phase events. Instead, mutations often move through the populations as temporal clusters (‘cohorts’) of functionally unrelated mutations, synchronously escaping drift and tracking tightly with one another through time. We quantify this temporal clustering of mutations in Fig. 2b, showing that it leads to a significant overrepresentation of time points at which either many or no mutations appeared, compared to the null expectation of mutations reaching detectable frequency at a constant rate ($P < 10^{-6}$).

Multiple mutations or cohorts of mutations are often present simultaneously, as is apparent in Fig. 1, and selective sweeps are often ‘nested’; that is, one sweep initiates before the preceding sweep has completed. Cohorts and nesting of mutations are forms of genetic hitchhiking, in which individual mutations are helped (or hindered) by the genetic background in which they happen to arise. This includes both hitchhiking of likely neutral synonymous mutations, as well as ‘quasi-hitchhiking’ of multiple beneficial mutations that act together as co-drivers. In addition, frequent interference between competing cohorts often leads to

¹Lewis-Sigler Institute for Integrative Genomics and Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA. ²Departments of Organismic and Evolutionary Biology and of Physics, and FAS Center for Systems Biology, Harvard University, Cambridge, Massachusetts 02138, USA. ³Departments of Biological Sciences and Chemistry & Biochemistry, Rowan University, Glassboro, New Jersey 08028, USA. ⁴The Genome Institute, Washington University School of Medicine, St. Louis, Missouri 63108, USA. [†]Present address: Department of Biological Sciences, Lehigh University, Bethlehem, Pennsylvania 18015, USA.

*These authors contributed equally to this work.

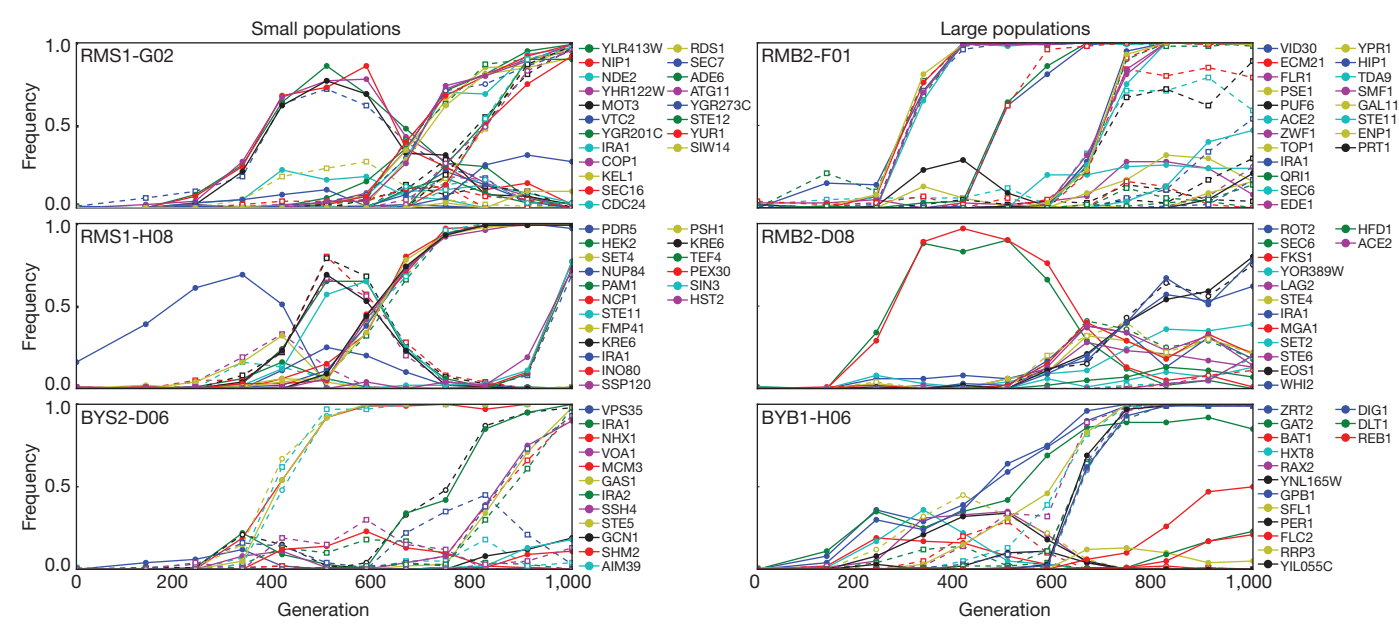


Figure 1 | The fates of individual spontaneously arising mutations. We show the frequency of all identified mutations through 1,000 generations in 6 of the 40 sequenced populations. Non-synonymous mutations are solid lines with solid circles, and synonymous and intergenic mutations are dotted lines with

open circles and squares, respectively. Populations in the left and right columns were evolved at small (10^5) and large (10^6) population sizes, respectively. We observe qualitatively similar patterns in the other populations (Supplementary Fig. 1).

the extinction of beneficial mutations even after they reach substantial frequency. Drawing from the full aggregate data set as well as individual ‘case study’ populations, we now show how this pervasive hitchhiking and interference strikes a balance between chance and determinism in governing evolutionary outcomes.

To investigate the repeatability of adaptation and identify those mutations that are driving adaptation, we looked for genes in which we observed mutations more often than expected by chance. Of the 995 nuclear mutations we identified, 723 fall within coding regions. If these mutations were distributed randomly over the 5,799 yeast genes, we expect only two genes with three or more mutations. Instead, we find

24 genes hit three or more times (Table 1, Supplementary Table 2 and Supplementary Fig. 2). This parallelism is at the gene level; mutations in different populations are different at the nucleotide level, with four exceptions (Methods). These 24 putative drivers represent approximately 0.6% of the yeast genome by size but account for 14% of the observed mutations, and are more likely to fix in the population (52/140, 37%) compared to all other nonsynonymous mutations (110/476, 23%, $P < 0.005$). Only 1 of the 141 mutations in these putative drivers is synonymous ($<1\%$), compared to 19% for the 472 mutations that fall in genes that are hit only once (Supplementary Table 3). Putative drivers are similarly depleted for missense mutations, and are enriched for nonsense and frameshift mutations (Supplementary Table 3). This mutational spectrum differs between functional categories of putative driver mutations. For genes in the mating pathway and

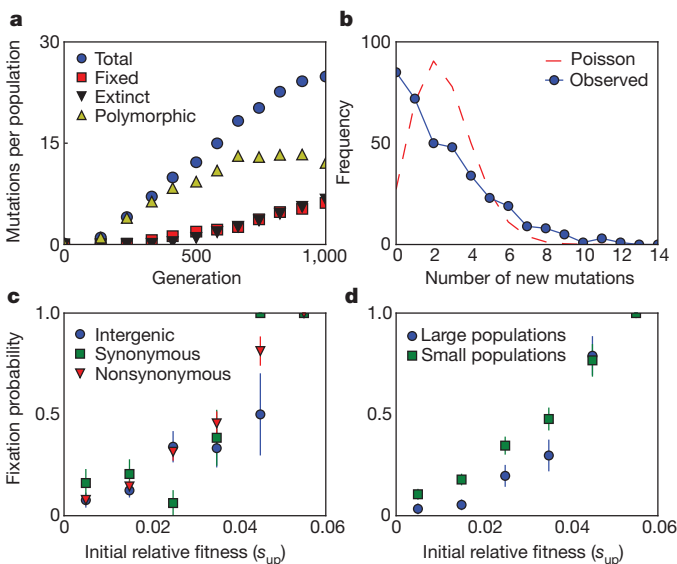


Figure 2 | Statistical analysis across 40 replicate populations. **a**, The per-population number of total mutations, fixed mutations, extinct mutations and mutations that are currently polymorphic over the course of the 1,000 generations. **b**, The distribution of the number of new mutations detected at each time point (solid blue line; see Methods for details) and a Poisson distribution with the same mean (dashed red line). **c**, **d**, Mutation fixation probability as a function of initial relative fitness. Data are mean \pm s.e.m.

Table 1 Repeatedly hit genes are putative drivers of adaptation			
Gene	Hits	Fixed	Biological process*
<i>IRA1</i>	21	10	Negative regulator of Ras
<i>ROT2</i>	11	2	Cell wall biogenesis
<i>YUR1</i>	11	5	Cell wall biogenesis
<i>ACE2</i>	9	4	Cytokinesis
<i>STE11</i>	9	1	Mating
<i>STE12</i>	9	2	Mating
<i>PDR5</i>	8	5	Multidrug transport
<i>WHI2</i>	7	2	General stress response
<i>STE4</i>	6	1	Mating
<i>IRA2</i>	5	3	Negative regulator of Ras
<i>KRE6</i>	4	1	Cell wall assembly
<i>SFL1</i>	4	1	Regulation of flocculation genes
<i>STE5</i>	4	3	Mating
<i>ANP1</i>	3	1	Protein glycosylation
<i>CNE1</i>	3	2	Protein folding
<i>GAS1</i>	3	3	Cell wall assembly
<i>GCN1</i>	3	1	Regulation of translation
<i>GPB1</i>	3	1	Negative regulator of Ras
<i>GPB2</i>	3	1	Negative regulator of Ras
<i>KEG1</i>	3	0	Cell wall assembly
<i>KRE5</i>	3	1	Cell wall assembly
<i>RPO31</i>	3	0	RNA polymerase III transcription
<i>SET4</i>	3	2	Unknown
<i>YJL171C</i>	3	0	Unknown

* Biological process was manually curated from the *Saccharomyces* Genome Database (<http://www.yeastgenome.org>).

negative regulators of Ras, we observe 14 missense, 8 nonsense, and 10 frameshift mutations, suggesting that selection at these loci is for loss of function (Supplementary Fig. 2). In contrast, all 13 mutations observed in cell-wall assembly genes are missense, suggesting that at these loci selection is for alteration or attenuation, not loss, of function (Supplementary Fig. 2).

This evidence argues that mutations in multi-hit genes provided strong fitness advantages that made them parallel adaptive solutions in multiple replicate populations (related arguments have been made in bacterial^{6,7} and viral²⁹ systems). However, the fate of each mutation also depends on random hitchhiking and interference effects, which increase variation in evolutionary outcomes. Even beneficial driver mutations must often quasi-hitchhike as co-drivers with others in a larger cohort if they are to succeed. For example, in population BYB1-G07, a mutation in *SPC3* began to sweep within the first 300 generations (Fig. 3), before a competing cohort appeared containing mutations in the multi-hit genes *WHI2* and *ROT2*. The *WHI2*–*ROT2* cohort rose in frequency at the expense of *SPC3*, until the *SPC3* genotype was partially rescued by a mutation in the multi-hit gene *YUR1*. Finally, a second and distinct mutation in *WHI2* appeared in the *SPC3*–*YUR1* background. This genotype fixed, forcing the *WHI2*–*ROT2* cohort to extinction. These dynamics illustrate how a balance between the fitness advantages of individual driver mutations and random hitchhiking and interference effects determines evolutionary outcomes.

Although the dynamics of any individual population are highly stochastic, a statistical analysis across replicate populations sheds light on the factors that determine the fate of each mutation. To this end, we measured the initial rate of increase in frequency of each mutation (Methods). We have previously²¹ referred to this ‘initial relative fitness’ as s_{up} . It measures the combined fitness effect of a mutation together with the genetic background in which it arose, relative to the average of all other genetic backgrounds currently in the population. The probability that a mutation fixes increases with s_{up} (Fig. 2c). Non-synonymous mutations tended to have higher s_{up} than synonymous mutations ($P < 0.05$) and non-synonymous mutations in multi-hit genes tended to have higher s_{up} than those in single-hit genes ($P < 0.02$), as we would expect if the former classes tend to confer a larger fitness advantage. However, given a particular value of s_{up} , all types of mutations were equally likely to succeed. In other words, a

weak or neutral mutation on a good background is just as likely to fix as a strongly beneficial mutation on a poor background; all that matters is the initial relative fitness of the mutation combined with the background in which it occurred.

In theory, population size could be predicted to either increase or decrease the patterns of reproducibility between replicate populations. Larger populations will sample more possible mutations, and thus favour the best genotypes in replicate populations^{13,16}. But larger populations also maintain more genetic variation, making each mutation more likely to be influenced by chance associations¹⁶. Our data make it possible to determine experimentally the influence of population size on the reproducibility of evolutionary outcomes. Of the 40 sequenced populations, 14 were evolved at a large (10^6) and 26 at a small (10^5) population size. We find that putative driver mutations are more commonly observed in large populations (Table 2, $P < 0.025$). However, given a particular value of s_{up} , a mutation is less likely to fix in a large population—that is, subsequent chance associations are more likely to interfere (Fig. 2d, $P < 10^{-5}$). Together, these results show that beneficial mutations occur more consistently in larger populations, but that each mutation has a more random fate once it has occurred.

To demonstrate how our system can be used to dissect the fitness effects of the individual mutations that underlie these dynamics, we chose for genetic dissection a population that displayed simple sequence-level dynamics. In BYS1-A08, two mutations (*ELO1* and *GAS1*) have fixed and a third (*STE12*) is on its way to fixation by generation 545 (Fig. 4a). Clones from this time point had gained, on average, a 4.3% fitness advantage relative to the ancestor. To determine how these three mutations contribute to fitness, we crossed three clones from generation 545 to the ancestor and isolated 80 haploid progeny. Each haploid was genotyped at these three loci and assayed for fitness, enabling us to quantify the fitness effect of each mutation individually and in combination. We find that mutations in both *GAS1* and *STE12* provide a selective advantage, while the mutation in *ELO1* is a neutral hitchhiker (Fig. 4b). Consistent with this, mutations in *GAS1* and *STE12* are observed in three and nine replicate populations, respectively (Table 1), but *ELO1* only once.

Our analysis has shown that the combination of experimental evolution and whole-genome whole-population sequencing over a dense timecourse is a powerful tool. Our data demonstrate the importance of pervasive hitchhiking and clonal interference among cohorts of mutations in determining the molecular dynamics of adaptation. Further work is needed to determine the mechanism underlying the formation of these cohorts. Interestingly, cohorts and genetic hitchhiking have been described in other systems, such as influenza evolution² and the somatic evolution of cancers³⁰, suggesting that these dynamics represent a general mode of adaptation. Our data also highlight the relatively small subset of genes that repeatedly provide driver mutations, suggesting a limited number of open pathways to substantially increased fitness. This work is a first step towards a complete understanding of the dynamics of adaptation under conditions where multiple beneficial mutations spread simultaneously, and illustrates the importance of both chance and selection in determining evolutionary outcomes.

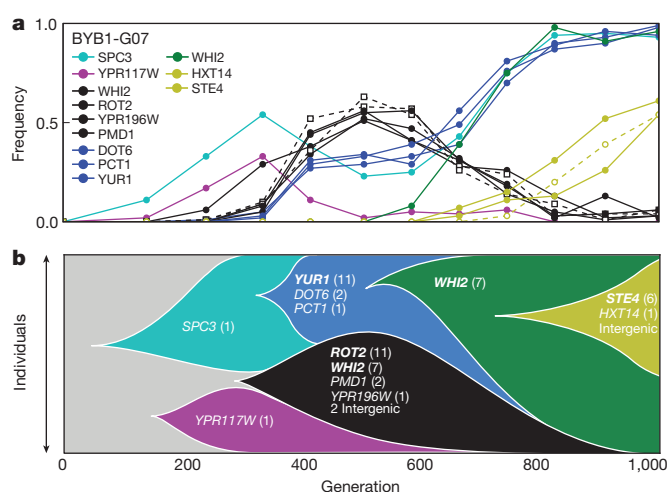


Figure 3 | The dynamics of sequence evolution in BYB1-G07. a, The trajectories of the 15 mutations that attain a frequency of at least 30%, hierarchically clustered into several distinct mutation ‘cohorts’, each of which is represented by a different colour (Methods). **b,** Muller diagram showing the dynamics of the six main cohorts in the population. The number of times a mutation was observed in a given gene across all 40 populations is indicated in parentheses. Mutations in genes observed in more than three replicate populations (Table 1) are indicated in bold.

Table 2 | Summary of the fates of nuclear mutations observed throughout the experiment

Class of mutation	All populations (40)			Small populations (26)			Large populations (14)		
	Total	No. fixed	Fixed (%)	Total	No. fixed	Fixed (%)	Total	No. fixed	Fixed (%)
All	995	246	25	703	191	27	292	55	19
Intergenic	272	58	21	207	48	23	65	10	15
Synonymous	107	26	24	77	21	27	30	5	17
Nonsynonymous	616	162	26	419	122	29	197	40	20
Hit 1 ×	381	86	23	273	65	24	108	21	19
Hit 2 ×	95	24	25	63	17	27	32	7	22
Hit ≥3 ×	140	52	37	83	40	48	57	12	21

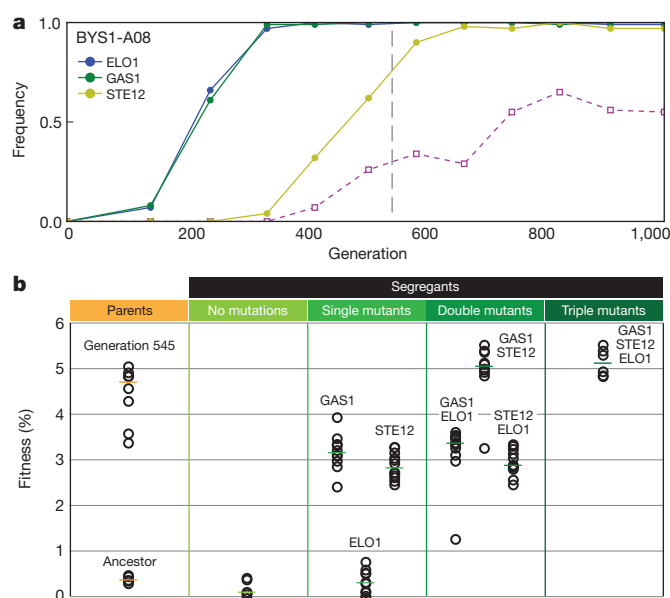


Figure 4 | Genetic dissection of BYS1-A08. **a**, The trajectories of observed mutations. **b**, We crossed evolved clones from generation 545 (dotted grey line in **a**) to the ancestor; shown here are the fitnesses and genotypes of parental clones and 80 haploid progeny.

METHODS SUMMARY

Whole-population DNA samples were sequenced using the Illumina HiSeq platform. Mutations were identified using Breseq (<http://www.barricklab.org/twiki/bin/view/Lab/ToolsBacterialGenomeResequencing;Pipeline1>), or BWA and FreeBayes (<http://www.bio-bwa.sourceforge.net> and Marth Laboratory, Boston College; Pipeline 2). Time-course information was used to distinguish real mutations from spurious calls due to sequencing or alignment error. Mutations were annotated to the yeast genome using NCBI-BLAST. Fitness of evolved or reconstructed clones was determined by competition against a fluorescently-labelled reference strain as described previously²¹.

Full Methods and any associated references are available in the online version of the paper.

Received 11 January; accepted 3 June 2013.

Published online 21 July 2013.

- Weinreich, D. M., Delaney, N. F., DePristo, M. A. & Hartl, D. L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006).
- Strelkova, N. & Lässig, M. Clonal interference in the evolution of influenza. *Genetics* **192**, 671–682 (2012).
- Levin, B. R. & Bull, J. J. Short-sighted evolution and the virulence of pathogenic microorganisms. *Trends Microbiol.* **2**, 76–81 (1994).
- Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481**, 306–313 (2012).
- Sprouffske, K., Merlo, L. M. F., Gerrish, P. J., Maley, C. C. & Sniegowski, P. D. Cancer in light of experimental evolution. *Curr. Biol.* **22**, R762–R771 (2012).
- Tenaillon, O. *et al.* The molecular diversity of adaptive convergence. *Science* **335**, 457–461 (2012).
- Woods, R., Schneider, D., Winkworth, C. L., Riley, M. A. & Lenski, R. E. Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **103**, 9107–9112 (2006).
- Saxer, G., Doebe, M. & Travisano, M. The repeatability of adaptive radiation during long-term experimental evolution of *Escherichia coli* in a multiple nutrient environment. *PLoS ONE* **5**, e14184 (2010).
- Atwood, K. C., Schneider, L. K. & Ryan, F. J. Periodic selection in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **37**, 146–155 (1951).

- Paquin, C. & Adams, J. Frequency of fixation of adaptive mutations is higher in evolving diploid than haploid yeast populations. *Nature* **302**, 495–500 (1983).
- Joseph, S. B. & Hall, D. W. Spontaneous mutations in diploid *Saccharomyces cerevisiae*: more beneficial than expected. *Genetics* **168**, 1817–1825 (2004).
- Perfeito, L., Fernandes, L., Mota, C. & Gordo, I. Adaptive mutations in bacteria: high rate and small effects. *Science* **317**, 813–815 (2007).
- Gerrish, P. J. & Lenski, R. The fate of competing beneficial mutations in an asexual population. *Genetica* **102–103**, 127–144 (1998).
- Desai, M. M. & Fisher, D. S. Beneficial mutation-selection balance and the effect of linkage on positive selection. *Genetics* **176**, 1759–1798 (2007).
- Rouzine, I. M., Wakeley, J. & Coffin, J. The solitary wave of asexual evolution. *Proc. Natl Acad. Sci. USA* **100**, 587–592 (2003).
- Good, B. H., Rouzine, I. M., Balick, D. J., Hallatschek, O. & Desai, M. M. The rate of adaptation and the distribution of fixed beneficial mutations in asexual populations. *Proc. Natl Acad. Sci. USA* **109**, 4950–4955 (2012).
- Schiffels, S., Szöllösi, G. J., Mustonen, V. & Lässig, M. Emergent neutrality in adaptive asexual evolution. *Genetics* **189**, 1361–1375 (2011).
- Desai, M. M., Fisher, D. S. & Murray, A. W. The speed of evolution and maintenance of variation in asexual populations. *Curr. Biol.* **17**, 385–394 (2007).
- de Visser, J. A. G. M., Zeyl, C. W., Gerrish, P. J., Blanchard, J. L. & Lenski, R. E. Diminishing returns from mutation supply rate in asexual populations. *Science* **283**, 404–406 (1999).
- Kao, K. C. & Sherlock, G. Molecular Characterization of clonal interference during adaptive evolution in asexual populations of *Saccharomyces cerevisiae*. *Nature Genet.* **40**, 1499–1504 (2008).
- Lang, G. I., Botstein, D. & Desai, M. M. Genetic variation and the fate of beneficial mutations in asexual populations. *Genetics* **188**, 647–661 (2011).
- Barrick, J. E. *et al.* Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* **461**, 1243–1247 (2009).
- Barrick, J. E. & Lenski, R. E. Genome-wide mutational diversity in an evolving population of *Escherichia coli*. *Cold Spring Harb. Symp. Quant. Biol.* **74**, 119–129 (2009).
- Dettman, J. R. *et al.* Evolutionary insight from whole-genome sequencing of experimentally evolved microbes. *Mol. Ecol.* **21**, 2058–2077 (2012).
- Gresham, D. *et al.* The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genet.* **4**, e1000303 (2008).
- Bollback, J. P. & Huelsenbeck, J. P. Clonal interference is alleviated by high mutation rates in large populations. *Mol. Biol. Evol.* **24**, 1397–1406 (2007).
- Betancourt, A. J. Genomewide patterns of substitution in adaptively evolving populations of the RNA bacteriophage MS2. *Genetics* **181**, 1535–1544 (2009).
- Miller, C. R., Joyce, P. & Wichman, H. A. Mutational effects and population dynamics during viral adaptation challenge current models. *Genetics* **187**, 185–202 (2011).
- Wichman, H. A., Badgett, M. R., Scott, L. A., Boulianne, C. M. & Bull, J. J. Different trajectories of parallel evolution during viral adaptation. *Science* **285**, 422–424 (1999).
- Nik-Zainal, S. *et al.* The life history of 21 breast cancers. *Cell* **149**, 994–1007 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the production team led by L. Fulton and R. Fulton at the Genome Institute at Washington University for sample management and data production, and E. Lobos for coordinating the project. We thank L. Parsons and J. Wiggins for assistance with data management, P. Gibney for assistance with sample preparation, and T. DeCoste for assistance with flow cytometry. We thank K. Kosheleva for discussions, and A. Murray, C. Marx, M. McDonald, G. Sherlock and D. Kvitek for comments on the manuscript. D.P.R. acknowledges support from an NSF Graduate Research Fellowship. D.B. acknowledges support from NIGMS Centers of Excellence grant GM071508 and NIH grant GM046406. M.M.D. acknowledges support from the James S. McDonnell Foundation, the Alfred P. Sloan Foundation, and the Harvard Milton Fund.

Author Contributions G.I.L., D.B. and M.M.D. designed the project; E.S. and G.M.W. generated the sequencing data; G.I.L., D.P.R., M.J.H. and M.M.D. analysed the sequencing data; G.I.L. performed the experiments; G.I.L., D.P.R., D.B. and M.M.D. wrote the paper. Co-senior authors, D.B. and M.M.D.

Author Information Genome sequence data have been deposited to GenBank under the BioProject identifier PRJNA205542. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.I.L. (glang@lehigh.edu) or M.M.D. (mmdesai@fas.harvard.edu).

METHODS

DNA sequencing. Cells were grown by inoculating 8 µl of each frozen population from our earlier experiment²¹ into 20 ml YPD (yeast extract, peptone, dextrose) + ampicillin (100 µg ml⁻¹) and tetracycline (25 µg ml⁻¹) and grown overnight to saturation. Cells were pelleted and washed once with water. Genomic DNA was prepared using a modified glass bead lysis method. Cells were resuspended in 400 µl of DNA extraction buffer (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris, pH 8.0, and 1 mM EDTA). To the resuspended cells, 600 µl of acid washed glass beads (425–600 µm, acid-washed; Sigma) and 400 µl of phenol:chloroform:isoamyl alcohol (25:24:1, Tris saturated) was added and the cells were mechanically lysed for 2.5 min using a bead beater. After centrifugation, the supernatant was removed and incubated at 37 °C with RNaseA for 1 h, followed by a second phenol:chloroform:isoamyl alcohol extraction. The aqueous supernatant was removed and genomic DNA was precipitated with ethanol and resuspended in water. Paired-end Illumina sequencing libraries of 500-bp fragments were prepared at The Genome Institute, Washington University School of Medicine, and the libraries were run on the Illumina HiSeq with average of 100-fold coverage.

Identifying mutations from raw sequencing data. We developed two independent methods for identifying mutations from the raw sequencing data and for distinguishing bona fide mutations from spurious calls that resulted from either sequencing or alignment errors by leveraging time course information. Both pipelines identified base-pair substitutions (BPS), small insertion and deletion mutations (InDel) and complex mutations involving both BPS and InDels. We note, however, that neither pipeline is well suited to identify certain types of mutations, such as copy number variation, inversions, or large insertions or deletions. Both pipelines produced similar results. The data presented in the paper were produced using Pipeline 1. Supplementary Table 1 reports the results of both pipelines.

In Pipeline 1, we used the software package Breseq (<http://www.barricklab.org/twiki/bin/view/Lab/ToolsBacterialGenomeResequencing>) to align Illumina reads and make initial polymorphism calls. We ran Breseq on each time point of each population independently and constructed a list of all mutations called in any time point. For each mutation, we used SAMTOOLS³¹ to calculate the frequency of reads supporting the mutation in all time points of the population where the mutation was called. We then applied a series of filters based on the frequency trajectories to eliminate false positives. Mutations that did not change frequency over the course of the entire experiment are likely to be sequencing or alignment errors. Therefore, we required the maximum frequency to be at least 0.1 greater than the minimum frequency. We also required the absolute difference between the maximum or minimum frequency and the frequency at generation zero to be at least 0.1. The frequency trajectories of real mutations are expected to be autocorrelated, whereas those of false positives should be uncorrelated from time point to time point. We rejected any mutation with an autocorrelation coefficient less than 0.2. Generation zero was not expected to contain any mutations. Therefore, we rejected any mutation detected by Breseq in generation zero of more than five populations. Also, for any mutation detected by Breseq in generation zero of more than two populations, we required the autocorrelation coefficient to be at least 0.5. Finally, for any mutation with a frequency greater than 0.01 in generation zero, we required the autocorrelation coefficient to be at least 0.35.

In Pipeline 2, for each population and for each time point, we aligned the raw reads to a SNP/Indel corrected W303 reference genome (reference available upon request) using BWA for Illumina version 1.2.2 (ref. 32) using default parameters (except 'Disallow insertion/deletion within [value] bp towards the end' set to 0 and 'Gap open penalty' set to 5). Mutations were called relative to the SNP/Indel corrected W303 reference genome using Freebayes version 0.8.9.a. (Marth Laboratory, Boston College) using default parameters (except 'Pooled' set to 'True', and 'Base alignment quality (BAQ) adjustment' set to 'True'). For each population we merged the 12 resulting .vcf files (one for each time point) using the 'vcf-merge' included in the VCFtools package (http://vcftools.sourceforge.net/perl_module.html). We wrote two perl scripts to analyse the resulting merged .vcf file (programs available upon request). The script 'allele_counts.pl' calculated the frequencies of mutant alleles for each time point in the series and 'composite_scores.pl' scored the trajectories of each mutation across the twelve time points based on six attributes: autocorrelation, area under the curve relative to time zero, minimum frequency, maximum frequency, max step (the largest difference in frequency in adjacent time points), and the number of called alternate alleles. We developed a heuristic composite score with which to rank the trajectories by their likelihood of being a bona fide mutation.

Any mutation called in either Pipeline was validated manually using the Integrative Genome Viewer^{33,34}.

Annotating mutations. For each mutation, we aligned the surrounding 2-kb region to the annotated s288c genome using NCBI-BLAST³⁵. We then used NCBI-BLAST's CDS feature option to identify the gene or intergenic region containing the mutation and the identity of any amino-acid changes.

All of the observed nuclear mutations represent unique alterations to the yeast genome with four exceptions: two cases of recurrent mutation at the same position and two instances of pre-existing mutations in the seed culture that reached detectable frequency during the evolution experiment. In *ROT2* and *STE12*, recurrent frameshift mutations were observed within homopolymeric runs of seven T's and eight G's, respectively. For *ROT2* all four occurrences of mutations in this homopolymeric run were T insertions. For *STE12*, two mutations were G insertions and two were G deletions. In addition to recurrent mutations, we observed two pre-existing mutations. In the initial evolution experiment, two nearly isogenic haploid ancestral strains (B and R) were used to seed approximately 300 populations each. Of the sequenced populations reported here, 30 are derived from the B progenitor and 10 from the R progenitor. We observed several occurrences where the same mutation was observed in multiple populations. The same single base-pair deletion in *IRA1* was observed in four populations derived from the B ancestor. In each case this allele was observed early and before the first selective sweep suggesting that this mutation was present at low frequency in the starting B population. In all 10 R populations, the same T to C substitution in *PDR5* was initially at 15% at Generation 0. This mutation quickly fixed in two populations, slowly fixed in another, rose to above 50% before going extinct in two and quickly went extinct in the other five (Supplementary Fig. 3).

Analysis of trajectories. To assess the relationship between fitness and fixation probability, we estimated s_{up} , the fitness of clones containing a given mutation relative to the mean fitness of the population when we first detected the mutation. For each mutation, we identified t_1 and t_2 , the first consecutive time points such that the frequency of the mutation at t_1 was greater than zero and the frequency at t_2 was greater than 0.1. We then calculated

$$s_{up} = \frac{1}{t_2 - t_1} \left(\ln \frac{f(t_2)}{1 - f(t_2)} - \ln \frac{f(t_1)}{1 - f(t_1)} \right)$$

where $f(t)$ is the frequency of the mutation at time t . This quantity estimates the combined effects of the focal mutation and the background it occurred on. For instance, a mutation conferring a 3% fitness advantage on a neutral background will have the same value of s_{up} as a neutral mutation occurring on a background that is 3% fitter than the population average.

Identifying mutation cohorts. The most notable feature of our results is that mutations often move through populations as temporal clusters of functionally unrelated mutations, tracking tightly with one another through time. We have termed these 'cohorts'. To empirically assign mutations to cohorts, we treated each frequency trajectory as a vector in twelve dimensions. We used the hierarchical clustering package in SciPy (<http://www.scipy.org>) to cluster the mutations in each population based on the Euclidean distance between frequency vectors. Because low-frequency mutations contain too little information for reliable clustering, we excluded mutations with maximum frequencies less than 0.3. We then flattened the hierarchies using a cutoff distance of 0.275.

Fitness assays and genetic dissection. Fitness assays were performed as described previously²¹. To measure the fitness of evolved clones from frozen stock, we struck to singles from population BYS1-A08 from generation 545. We selected seven single colonies at random and measured their fitness relative to an mCherry-expressing reference strain. The experimental and reference strains were grown separately in 96-well plates, then mixed 50:50 and propagated by diluting 1:1,024 every 24 h. At generations 10, 20, 30 and 40, we transferred 4 µl of saturated culture into 100 µl of cold PBST and the ratio of nonfluorescent (experimental) and mCherry-positive (reference) cells was determined by flow cytometry using an LSRII flow cytometer (BD Biosciences) counting 50,000 total cells for each sample. The fitness difference between the experimental and reference strain was calculated as the rate of the change in the \ln ratio of experimental to reference versus generations³⁶. To determine fitness effects of the three evolved mutations in BYS1-A08 (*GAS1*, *ELO1*, *STE12*), we chose three of the seven clones and backcrossed them to a MAT α version of the ancestral strain. From these three diploids we sporulated and selected 80 haploid MAT α segregants. Each segregant was genotyped by SNP-specific PCR using the following primers: *GAS1_Forward* (5'-TTTTCGTGCGCGCAAACGTGG-3'), *GAS1_WT_Reverse* (5'-ATTGGAAGAGTAGCCAACTG-3'), *GAS1_Mutant_Reverse* (5'-ATTGGAAGAGTAGCCAACTA-3'), *ELO1_Forward* (5'-AACACAACAAATCGCAAGCC-3'), *ELO1_WT_Reverse* (5'-TAACCAACCAATTGATTATA-3'), *ELO1_Mutant_Reverse* (5'-TAACCAACCAATTGATTATG-3'), *STE12_Reverse* (5'-TGAGCAGAACTCTTCGTCACC-3'), *STE12_WT_Forward* (5'-AATCTCACAACCTCTGGCCAG-3'), and *STE12_Mutant_Forward* (5'-AAATCTCACAACCTCTGCCAA-3'). The fitness of each of the haploid segregants was measured relative to the mCherry-expressing reference strain as described above.

31. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
32. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
33. Robinson, J. T. *et al.* Integrative genomics viewer. *Nature Biotechnol.* **29**, 24–26 (2011).
34. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
35. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
36. Hartl, D. *A Primer of Population Genetics*. (Sinauer Associates, 2000).

Prolonged dopamine signalling in striatum signals proximity and value of distant rewards

Mark W. Howe¹, Patrick L. Tierney¹, Stefan G. Sandberg², Paul E. M. Phillips² & Ann M. Graybiel¹

Predictions about future rewarding events have a powerful influence on behaviour. The phasic spike activity of dopamine-containing neurons, and corresponding dopamine transients in the striatum, are thought to underlie these predictions, encoding positive and negative reward prediction errors^{1–5}. However, many behaviours are directed towards distant goals, for which transient signals may fail to provide sustained drive. Here we report an extended mode of reward-predictive dopamine signalling in the striatum that emerged as rats moved towards distant goals. These dopamine signals, which were detected with fast-scan cyclic voltammetry (FSCV), gradually increased or—in rare instances—decreased as the animals navigated mazes to reach remote rewards, rather than having phasic or steady tonic profiles. These dopamine increases (ramps) scaled flexibly with both the distance and size of the rewards. During learning, these dopamine signals showed spatial preferences for goals in different locations and readily changed in magnitude to reflect changing values of the distant rewards. Such prolonged dopamine signalling could provide sustained motivational drive, a control mechanism that may be important for normal behaviour and that can be impaired in a range of neurologic and neuropsychiatric disorders.

The spike activity patterns of midbrain dopamine-containing neurons signal unexpected and salient cues and outcomes^{1–4,6,7}, and the dynamics of these phasic neural signals have been found to follow closely the principles of reinforcement learning theory^{3–6}. In accordance with this view, selective genetic manipulation of the phasic firing of dopamine neurons alters some forms of learning and cue-guided movements^{8,9}. Episodes of transient dopamine release in the ventral striatum have been detected with FSCV, and these also occur in response to primary rewards and, after learning, to cues predicting upcoming rewards^{10–13}. Thus, dopamine transients in the striatum share many features of the phasic spike activity of midbrain dopamine neurons.

Classic studies of such dopamine transients have focused on Pavlovian and instrumental lever-press tasks, in which rewards were within arm's reach^{1–3,10–13}. However, in many real-life situations, animals must move over large distances to reach their goals. These behaviours require ongoing motivational levels to be adjusted flexibly according to changing environmental conditions. The importance of such control of ongoing motivation is reflected in the severe impairments suffered in dopamine deficiency disorders, including Parkinson's disease. In addition, in pioneering experimental studies, dopamine signalling has been implicated in controlling levels of effort, vigour and motivation during the pursuit of goals in maze tasks^{14–17}. It has been unclear how phasic dopamine signalling alone could account for persistent motivational states¹⁸. We adapted chronic FSCV to enable prolonged measurement of real-time striatal dopamine release as animals learned to navigate towards spatially distant rewards.

We measured dopamine levels in the dorsolateral striatum (DLS) and ventromedial striatum (VMS) (Extended Data Figs 1 and 2, and Methods) as rats navigated mazes of different sizes and shapes to retrieve rewards (Figs 1–4, and Methods). The rats were trained first on an associative T-maze task to run and to turn right or left as instructed

by tones to receive a chocolate milk reward at the indicated end-arms¹⁹ ($n = 9$, Figs 1, 2 and 4). Unexpectedly, instead of mainly finding isolated dopamine transients at the initial cue or at goal-reaching, we primarily found gradual increases in the dopamine signals that began at the onset of the trial and ended after goal-reaching (Fig. 1a, b). These 'ramping' dopamine responses, identified in session averages by linear regression (Pearson's $R > 0.5$, $P < 0.01$), were most common in the VMS (75% of sessions) but were also present at DLS recording sites (42% of sessions). They were evident both in single trials (Fig. 1a–c) and in population averages (Fig. 1e, f, and Extended Data Figs 2g, h and 3), bore no clear relationship to run speed within or across trials (Fig. 1d), and matched, in electrochemical profile, dopamine release evoked by tonic electrical stimulation *in vivo* (Extended Data Fig. 2i, j). Before goal-reaching, the ramps had similar amplitudes in correct (65% overall) and incorrect trials (Fig. 1e, f). After goal-reaching, the signals were significantly larger in correct trials, particularly in the VMS (paired t -test, $P = 0.01$, Fig. 1e, f). Notably, a subset of the session-averaged signals in the DLS (22%, 58 out of 262; 7 probes in 5 rats) showed sustained inhibition up to goal-reaching (Extended Data Fig. 3). Such negative signals were rare in the VMS (5%, 15 out of 300 recordings), suggesting that ramping dopamine signals in the DLS, but not VMS, exhibit heterogeneity in polarity.

We identified isolated phasic transients at warning click indicating trial start and after goal-reaching. These were clearly distinct from the slower ramping responses in approximately 10% of single trials (Extended Data Fig. 4) but were often superimposed on the ramping signals, indicating that the signals recorded could include combinations of transient increases after warning click, slower ramps to goal-reaching, and transient increases after goal-reaching (Extended Data Fig. 4d). The peak magnitudes of the dopamine ramps were comparable to, or slightly smaller than, those of isolated phasic dopamine signals recorded here (Extended Data Fig. 4) and in other studies^{11,12}, and they were correlated with the peak magnitudes of free-reward evoked dopamine measured on the same probes (Pearson's $R = 0.45$, $P < 0.001$, Extended Data Fig. 5), indicating that the ramping signals could be subject to similar regulatory mechanisms and display similar anatomic heterogeneity as classical phasic reward-evoked dopamine signals.

We took advantage of the trial-to-trial variability in the rats' run times (Fig. 2a) to determine whether ramping dopamine release reflected elapsed time or reward proximity, or whether the ramps reflected sums of multiple, accumulated transients to fixed maze cues²⁰. If the dopamine ramps tracked elapsed time, peak dopamine values should have scaled directly with trial time (same slope, different peak height; Fig. 2b, f). If the ramping reflected distance or spatial location relative to goal-reaching (proximity), peak dopamine levels should have been equivalent for shorter and longer trials (different slope, same peak height; Fig. 2c, f). If the ramps were generated by summation of multiple transients, then for characteristic transient dynamics, the signals should have tended to peak at lower values for longer runs than for shorter runs (different slope, different peak height; Extended Data Fig. 6a, b, and Supplementary Discussion). The measured peak dopamine values at goal-reaching were nearly equivalent for short and long

¹McGovern Institute for Brain Research and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ²Department of Psychiatry and Behavioral Sciences, and Department of Pharmacology, University of Washington, Seattle, Washington 98195, USA.

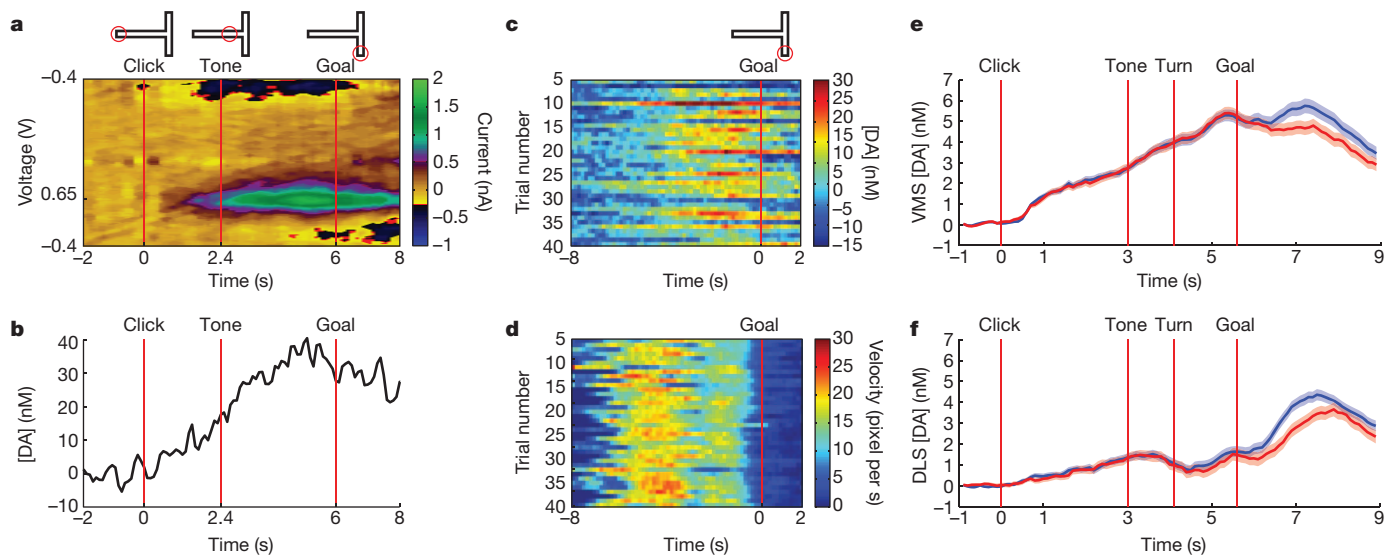


Figure 1 | Ramping striatal dopamine signals occur during maze runs. a, b, Baseline subtracted current (a) and dopamine concentration ([DA], b) measured by FSCV in VMS during a single T-maze trial. c, d, Trial-by-trial changes in dopamine concentration (c) and velocity (d) relative to

goal-reaching. e, f, Dopamine concentration (mean \pm s.e.m.) for VMS (e, $n = 300$ session-averaged recordings from 18 probes across 214 sessions) and for DLS (f, $n = 262$, 13 probes) for correct (blue) and incorrect (red) trials, averaged over all 40 trial sessions.

trials (Fig. 2e), and were not correlated with trial length (Fig. 2d–f) or with run velocity or acceleration (Extended Data Fig. 6e, f). Moreover, on trials in which rats paused mid-run, the signals remained sustained (or dipped slightly) and resembled the actual proximity to reward (Extended Data Fig. 7). These observations indicated that the ramping signals could represent a novel form of dopamine signalling that provides a continuous estimate of the animal's spatial proximity to distant rewards (Fig. 2, Extended Data Fig. 6, and Supplementary Discussion).

Given that phasic responses of dopamine-containing neurons can reflect the relative value of stimuli²¹, we asked, in a subset of rats, whether the ramping dopamine signals could also be modulated by the size of the delivered rewards (Methods). We used mazes with T, M or S configurations and different total lengths (Fig. 3, and Extended Data Fig. 8). We required the animals to run towards one or the other end of the maze and varied the rewards available at the alternate goal regions. With all three mazes, dopamine ramping became strongly biased towards the goal with the larger reward (Fig. 3, and Extended Data Fig. 8). Run speed was slightly higher for the high-reward maze arms (Fig. 3i, k), but these small differences were unlikely to account

fully for the large differences in the dopamine signals recorded. When we then reversed the locations of the small and large rewards, the ramping signals also shifted, across sessions or just a few trials, to favour the new high-value maze arm (Fig. 3, and Extended Data Fig. 8). These bias effects were statistically significant for each experimental paradigm (Extended Data Fig. 8h–j, Mann–Whitney U-test, $P < 0.05$) and across all rats (Fig. 3d, $n = 4$, Mann–Whitney U-test, $P = 0.02$).

In the M-maze, the ramps became extended to cover the longer end-arm distances to goal-reaching, and critically, peaked at nearly the same level before goal-reaching as did the ramping signals recorded in the T-maze, despite the longer distance travelled (Fig. 3e). This result suggested that the ramping dopamine signals do not signal reward proximity in absolute terms but, instead, scale with the path distance to a fixed level that depends on the relative reward value.

To determine whether such value-related differences in the ramping dopamine signals would occur when the actions to reach the distant goal sites were equivalent, we used the S-shaped maze. The ramping signals were larger for the run trajectories leading to the larger rewards (Fig. 3c, j, and Extended Data Fig. 9), despite the fact that the sequence

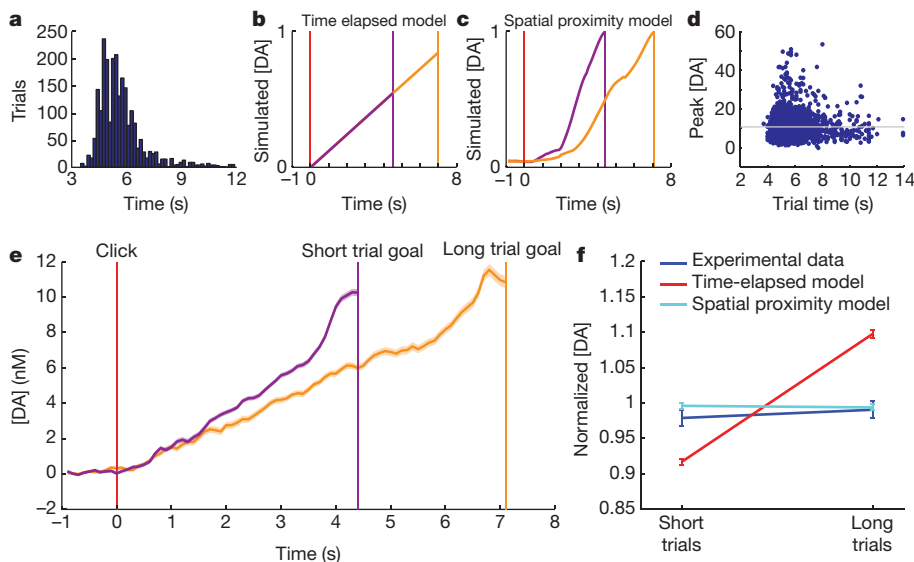


Figure 2 | Ramping dopamine signals proximity to distant rewards. a, Distribution of trial times (from warning click to goal-reaching, $n = 3,933$ trials). b, c, Dopamine release modelled as a function of time elapsed since maze-running onset (b) and as a function of spatial proximity to visited goal (c) for short (purple) and long (orange) trials (see Methods). Vertical lines indicate trial start (red) and end (purple and orange) times. d, Peak dopamine concentration versus trial time for all ramping trials ($n = 2,273$, Pearson's $R = 0.0004$, $P = 0.98$). e, Experimentally recorded dopamine release (mean \pm s.e.m.) in short ($n = 327$, purple) and long ($n = 423$, orange) trials. Dopamine peaks at equivalent levels, as in the proximity model in c. f, Normalized peak dopamine levels (mean \pm s.e.m.) predicted by time-elapsed (red) and proximity (light blue) models, and measured experimental data (dark blue).

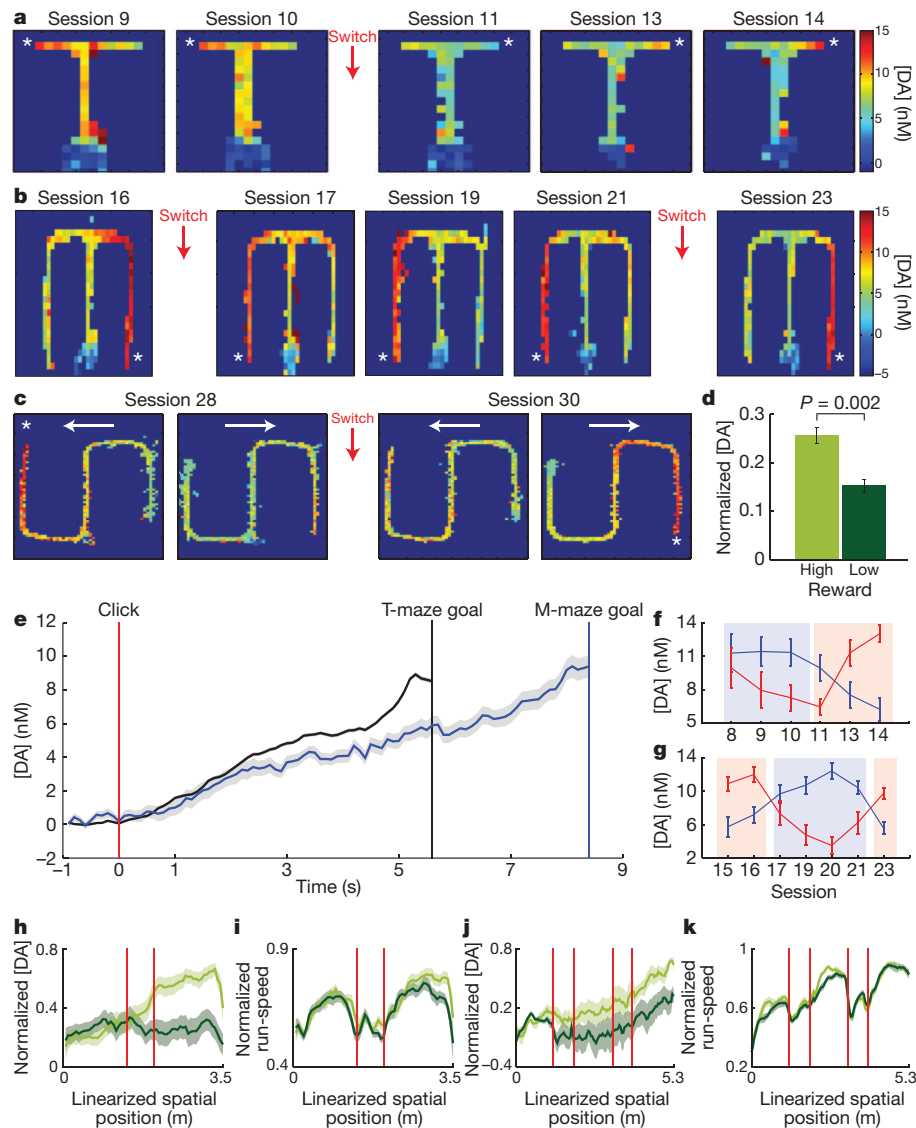


Figure 3 | Dopamine ramping is sensitive to reward magnitude.

a, b, Average dopamine signals from a VMS probe, for consecutive T-maze (**a**) and M-maze (**b**) sessions with asymmetric rewards. Asterisks indicate the goal with the larger reward; red arrows (and Switch) indicate reversal of reward amounts. **c**, Dopamine signals from a different rat running in the S-maze. White arrows indicate run direction. **d**, Average (\pm s.e.m.) peak dopamine across all value experiments ($n = 4$ rats). **e**, Average (\pm s.e.m.) VMS dopamine during T-maze ($n = 44$ sessions in 3 rats, black) and M-maze ($n = 17$, blue)

sessions in the same rats. **f, g**, Average (\pm s.e.m.) peak dopamine signals for the sessions plotted in **a** (**f**) and **b** (**g**) for trials to left (blue) and right (red) goals. Shading indicates arm with larger reward. **h, i**, Average normalized dopamine (**h**) and running speed (**i**) for runs to high (light green) and low (dark green) reward goals in the M-maze. Vertical lines indicate turns. **j, k**, Average normalized dopamine (**j**) and running speed (**k**) in the S-maze ($n = 9$ sessions in 2 rats), plotted as in **h** and **i**.

of turns and the lengths of the runs needed to reach the larger and smaller rewards were equivalent for both trajectories ($n = 2$ rats, 4 and 5 sessions per rat, Fig. 3c, j, k, and Extended Data Figs 8 and 9).

In rats performing the free-choice associative version of the T-maze task, robust dopamine signal biases existed in approximately 20% of sessions (Mann–Whitney U-test, $P < 0.05$) and significantly more often than chance overall (Z-test, $P < 0.00001$ versus bootstrapped variances; Methods and Fig. 4a, b, d). These biases were maintained across consecutive training sessions for individual animals (Fig. 4b), did not relate to run-speed biases (Fig. 4c, and Extended Data Fig. 10c) or recording hemisphere (Fig. 4d, and Extended Data Fig. 10a, b, f), and, notably, emerged gradually over days as performance improved and training progressed (Fig. 4e, f). Although not obviously related to imbalances in maze cues or differences in left–right performance, they displayed a weak association with right end-arm choice biases that developed late in training (Extended Data Fig. 10d, e, g). Thus, end-arm biases in the ramping dopamine signals could develop even in the absence

of experimentally imposed discrepancies in value, possibly reflecting developing internal value estimates (Supplementary Discussion).

Ramping spike-firing has been recorded for putative midbrain dopamine neurons in head-fixed primates under conditions of reward uncertainty²² and for nigral non-dopamine-containing neurons¹. We asked whether the magnitudes of the ramping dopamine signals that we recorded in the striatum changed as performance improved on the free-choice associative T-maze task (Fig. 4e). They did not (Pearson's $R = -0.08$, $P = 0.19$; Extended Data Fig. 10h–j), suggesting that uncertainty about reward probability was unlikely to have controlled the magnitude of the ramping signals²² (Supplementary Discussion).

Classic studies of dopamine neuron firing and striatal dopamine release have largely focused on transient responses associated with unpredicted rewards and reward-predictive cues. Here we demonstrate that, in addition to such transient dopamine responses, prolonged dopamine release in the striatum can occur, changing slowly as animals approach distant rewards during spatial navigation. These

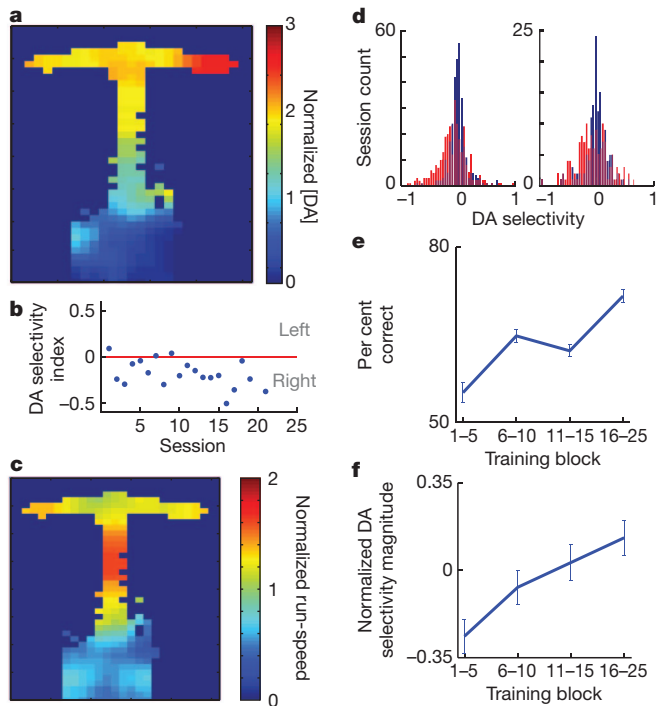


Figure 4 | Ramping dopamine selectivity can emerge with training without experimentally imposed reward discrepancies. **a**, Average normalized dopamine at a VMS site as a function of maze location ($n = 19$ sessions). **b**, Dopamine selectivity indices (Methods) for all individual sessions averaged in **a**. **c**, Average running speed for sessions in **a**. **d**, Selectivity indices for all VMS (left) and DLS (right) recordings (red) compared to shuffled data (blue) for all rats ($n = 9$). **e**, **f**, Average per cent correct performance (**e**) and average Z-score normalized dopamine selectivity (**f**) across training blocks. Error bars, s.e.m.

dopamine signals seem to represent the relative spatial proximity of valued goals, perhaps reflecting reward expectation²³. It remains unclear whether these signals represent goal proximity on the basis of environmental cues, effort, or internally scaled estimates of distance. However, the brain possesses mechanisms for representing both allocentric spatial context and relative distance from landmarks²⁴, which could, in principle, be integrated with dopaminergic signalling to produce such extended dopamine signals.

Transient dopaminergic responses to learned reward-predictive cues have been proposed to initiate motivated behaviours^{25,26}, but with this mode of signalling alone, it is difficult to account for how dopamine acts to maintain and direct motivational resources during prolonged behaviours (Supplementary Discussion). The ramping dopamine signals that we describe here, providing continuous estimates of how close rewards are to being reached, and weighted by the relative values of the rewards when options are available, seem ideally suited to maintain and direct such extended energy and motivation.

METHODS SUMMARY

Male Long Evans rats ($n = 9$) were deeply anaesthetized and were surgically implanted with headstages carrying voltammetry microsensor placed in the VMS (anteroposterior ± 1.5 mm, mediolateral ± 2.1 mm, dorsoventral 6–7 mm) and DLS (anteroposterior ± 0.5 mm, mediolateral ± 3.5 mm, dorsoventral 3.5–4.0 mm) bilaterally ($n = 1$ rat) or in the left ($n = 5$) or right ($n = 3$) hemisphere. Triangular voltage sweeps (-0.4 V to 1.3 V) were applied at 10 Hz to the carbon fibre microsensor probes²⁷ relative to an Ag/AgCl reference electrode implanted in the neocortex. Current changes due to dopamine oxidation–reduction reactions were verified by principal component regression²⁸ using a training set of current profiles from evoked dopamine release and pH changes recorded in 5 rats in response to electrical stimulation of the medial forebrain bundle (MFB). Training on the T-maze task¹⁹ (Figs 1, 2 and 4) and chronic voltammetry recording began 4 weeks after surgery, and continued for 15 to 35 days for each rat. The position

of each rat was monitored continually by tracking software as the rat navigated the maze (Neuralynx), and the positions were used to trigger auditory cues and reward pumps when the animal entered specific maze locations. After T-maze training, a subset of rats ($n = 3$) was trained in an extended M-maze. Rats were required either to continue performing the associative tone-cued task (M31) or to visit one of the two end-arms on each trial by blockade of the other end-arm (M36 and M47). Another subset of rats ($n = 2$) underwent training on an S-maze version of the task in which they only had to run back and forth from end to end to retrieve rewards. Voltammetry data were analysed with in-house MATLAB codes (Mathworks), and probe positions were verified by standard histology¹⁹.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 January; accepted 17 July 2013.

Published online 4 August 2013.

- Cohen, J. Y., Haesler, S., Vong, L., Lowell, B. B. & Uchida, N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* **482**, 85–88 (2012).
- Matsumoto, M. & Hikosaka, O. Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature* **459**, 837–841 (2009).
- Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
- Waelti, P., Dickinson, A. & Schultz, W. Dopamine responses comply with basic assumptions of formal learning theory. *Nature* **412**, 43–48 (2001).
- Bayer, H. M. & Glimcher, P. W. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* **47**, 129–141 (2005).
- Schultz, W. Getting formal with dopamine and reward. *Neuron* **36**, 241–263 (2002).
- Redgrave, P. & Gurney, K. The short-latency dopamine signal: a role in discovering novel actions? *Nature Rev. Neurosci.* **7**, 967–975 (2006).
- Zweifel, L. S. *et al.* Disruption of NMDAR-dependent burst firing by dopamine neurons provides selective assessment of phasic dopamine-dependent behavior. *Proc. Natl Acad. Sci. USA* **106**, 7281–7288 (2009).
- Wang, L. P. *et al.* NMDA receptors in dopaminergic neurons are crucial for habit learning. *Neuron* **72**, 1055–1066 (2011).
- Day, J. J., Jones, J. L., Wightman, R. M. & Carelli, R. M. Phasic nucleus accumbens dopamine release encodes effort- and delay-related costs. *Biol. Psychiatry* **68**, 306–309 (2010).
- Day, J. J., Roitman, M. F., Wightman, R. M. & Carelli, R. M. Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nature Neurosci.* **10**, 1020–1028 (2007).
- Gan, J. O., Walton, M. E. & Phillips, P. E. Dissociable cost and benefit encoding of future rewards by mesolimbic dopamine. *Nature Neurosci.* **13**, 25–27 (2010).
- Phillips, P. E., Stuber, G. D., Heien, M. L., Wightman, R. M. & Carelli, R. M. Subsecond dopamine release promotes cocaine seeking. *Nature* **422**, 614–618 (2003).
- Braun, A. A., Graham, D. L., Schaefer, T. L., Vorhees, C. V. & Williams, M. T. Dorsal striatal dopamine depletion impairs both allocentric and egocentric navigation in rats. *Neurobiol. Learn. Mem.* **97**, 402–408 (2012).
- Salamone, J. D., Correa, M., Farrar, A. & Mingote, S. M. Effort-related functions of nucleus accumbens dopamine and associated forebrain circuits. *Psychopharmacology (Berl.)* **191**, 461–482 (2007).
- Whishaw, I. Q. & Dunnett, S. B. Dopamine depletion, stimulation or blockade in the rat disrupts spatial navigation and locomotion dependent upon beacon or distal cues. *Behav. Brain Res.* **18**, 11–29 (1985).
- Salamone, J. D. & Correa, M. The mysterious motivational functions of mesolimbic dopamine. *Neuron* **76**, 470–485 (2012).
- Niv, Y., Daw, N. D., Joel, D. & Dayan, P. Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology (Berl.)* **191**, 507–520 (2007).
- Barnes, T. D., Kubota, Y., Hu, D., Jin, D. Z. & Graybiel, A. M. Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature* **437**, 1158–1161 (2005).
- Niv, Y., Duff, M. O. & Dayan, P. Dopamine, uncertainty and TD learning. *Behav. Brain Funct.* **1**, 6 (2005).
- Tobler, P. N., Fiorillo, C. D. & Schultz, W. Adaptive coding of reward value by dopamine neurons. *Science* **307**, 1642–1645 (2005).
- Fiorillo, C. D., Tobler, P. N. & Schultz, W. Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* **299**, 1898–1902 (2003).
- Hikosaka, O., Sakamoto, M. & Usui, S. Functional properties of monkey caudate neurons. III. Activities related to expectation of target and reward. *J. Neurophysiol.* **61**, 814–832 (1989).
- Derdikman, D. & Moser, E. I. A manifold of spatial maps in the brain. *Trends Cogn. Sci.* **14**, 561–569 (2010).
- Flagel, S. B. *et al.* A selective role for dopamine in stimulus-reward learning. *Nature* **469**, 53–57 (2011).
- Berridge, K. C. The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology (Berl.)* **191**, 391–431 (2007).

27. Clark, J. J. *et al.* Chronic microensors for longitudinal, subsecond dopamine detection in behaving animals. *Nature Methods* **7**, 126–129 (2010).
28. Keithley, R. B., Heien, M. L. & Wightman, R. M. Multivariate concentration determination using principal component regression with residual analysis. *Trends Analyt. Chem.* **28**, 1127–1136 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank G. Telian, L. Li, T. Lechina, S. Ng-Evans and N. Hollon for help, and M. J. Kim, K. S. Smith, T. J. Sejnowski, L. G. Gibb and Y. Kubota for their comments. This work was supported by US National Institutes of Health (NIH) grant R01 MH060379 (A.M.G.), National Parkinson Foundation (A.M.G.), CHDI Foundation

grant A-5552, the Stanley H. and Sheila G. Sydney Fund (A.M.G.), a Mark Gorenberg fellowship (M.W.H.), and NIH grants R01 DA027858 and R01 MH079292 (P.E.M.P.).

Author Contributions M.W.H. and A.M.G. designed the experiments, analysed the data, and wrote the manuscript. M.W.H. conducted the experiments. P.L.T. helped to develop the FSCV chronic recording setup. S.G.S. and P.E.M.P. provided technical training on FSCV and advice on data analysis. All authors contributed to discussion and interpretation of the findings.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.M.G. (graybiel@mit.edu).

METHODS

All experimental procedures were approved by the Committee on Animal Care at the Massachusetts Institute of Technology and were in accordance with the US National Research Council Guide for the Care and Use of Laboratory Animals. Sample sizes were chosen based on signal variability estimates from other published studies using FSCV.

Implant procedures. Implantations were performed on deeply anaesthetized fully mature male Long Evans rats ($n = 9$; 380 to 450 g), in sterile conditions according to approved surgical procedures¹⁹ with headstages carrying 1 to 3 independently movable voltammetry probes targeting the DLS (anteroposterior +0.5 mm, mediolateral ± 3.5 mm, dorsoventral 3.5–4.0 mm), of the right ($n = 3$) or left ($n = 5$) hemisphere, or the DLS bilaterally ($n = 1$), 1 to 3 probes targeting the VMS of the same hemispheres (anteroposterior +1.5 mm, mediolateral ± 2.1 mm, dorsoventral 6–7 mm), and a unilateral Ag/AgCl reference electrode in the posterior cortex (anteroposterior –2.3 mm, mediolateral ± 3.5 mm, dorsoventral ~ 0.5 mm). Five rats that underwent maze training and three additional rats for acute stimulation experiments were implanted with tungsten bipolar stimulation electrodes (FHC) straddling the ipsilateral MFB (anteroposterior –4.6 mm, mediolateral ± 1.3 mm, dorsoventral 7–8 mm) to verify striatal dopamine release (see below).

Behavioural training. All behavioural training was conducted on a custom built 'grid maze' with fully reconfigurable tracks and walls. Training on an associative T-maze task with auditory instruction cues (Figs 1, 2, and 4)¹⁹ began 4 weeks after implantation. Voltammetric recordings began when rats learned to run smoothly down the track to retrieve reward. Early sessions with sporadic maze behaviour, such as wall rearing and sluggish initiation of maze running, were discarded. Daily behavioural sessions consisted of 40 trials. Trials began with a warning click, followed 0.5 s later by the lowering of a swinging gate, allowing the rat to run down the maze. Half-way down the long arm, a tone was triggered (1 or 8 kHz), indicating which end-arm to visit in order to retrieve chocolate milk reward (0.3 ml) delivered through automated syringe pumps (Pump Systems) upon the rat's arrival. The spatial position of each rat was monitored continually by video tracking (Neuralynx). Tone delivery and syringe pumps were controlled by in-house behavioural software written in MATLAB (Mathworks). After 15 to 35 T-maze sessions per rat, a subset of rats ($n = 3$) received 17 training sessions (4 to 6 sessions each) on the M-maze task in which the end-arms of the T-maze were extended (Fig. 3). These rats received a larger amount of reward (0.4 ml) at one goal site than at the other (0.1 ml for 2 rats and 0.2 ml for 1 rat). After 2 to 3 sessions with a given set of spatial reward contingencies, the reward amounts at the two goals were reversed. One rat (M31) was required to make turn choices in response to tones as in the previous T-maze task, whereas the other two rats (M36 and M47) were directed pseudo-randomly to one end-arm of the maze on each trial by removing the track to the opposite arm (20 trials to each arm) without tone presentation. Two rats were trained on the S-maze task (Fig. 3). These rats were required simply to run back and forth to retrieve a large volume of chocolate milk (0.4 ml) at one goal and a small volume (0.1 ml) at the other goal. Consecutive visits to the same reward site did not trigger the reward pumps.

Voltammetry data acquisition and analysis. Waveform generation and data acquisition for voltammetry recordings were done with two PCI data acquisition cards and software written in LabVIEW (National Instruments). Triangular voltage waveforms were applied to chronically implanted carbon fibre electrodes, relative to the reference electrode, at 10 Hz. Electrodes were held at –0.4 V between scans, and were ramped to 1.3 V and back to –0.4 V during each scan²⁷. Current produced by redox reactions was recorded during voltage scans.

We compiled a library of current versus applied voltage templates for dopamine and pH changes of varying magnitudes by stimulating the MFB (60 Hz, 24 pulses, 300 μ A) to induce dopamine release in the striatum in 5 rats maintained under isoflurane anaesthesia. We used these templates from all 5 rats as a training set to perform chemometric analysis²⁸ on voltammetry measurements obtained during behaviour with in-house MATLAB software. This procedure allowed us to distinguish changes in current due to dopamine release from changes due to pH or to other electroactive substances²⁸. In a separate set of rats, we stimulated the MFB (10 Hz, 60 pulses, 100–120 μ A) under urethane anaesthesia to mimic the slower, low amplitude ramping signals that we observed in behaving animals (Extended Data Fig. 2). Current changes were converted to estimated dopamine concentration by using calibration factors obtained from *in vitro* measurements of fixed dopamine concentrations. Behavioural video tracking was synchronized with voltammetry recordings by marker transistor–transistor logic signals sent to the voltammetry data acquisition system.

For each trial, voltammetry data were normalized by subtracting average background current at each potential measured during the 1-s baseline period before warning click. Session averaged traces (Figs 1 and 3, and Extended Data Figs 3 and 10) were computed by averaging the dopamine signals recorded in a single session across 40 trials, and then averaging these traces to obtain global averages across all

rats and electrodes. Each session-averaged trace (one from each probe from each session) was considered as an independent measurement for computing s.e.m. Concatenation of dopamine and proximity signals (see below) was performed by scaling the peri-event windows using the median inter-event intervals between consecutive events across all trials (Figs 1–3, and Extended Data Figs 3, 4 and 10). Traces between two consecutive events were plotted by taking data from each event to half of the median inter-event interval. Maze arm selectivity (Fig. 4 and Extended Data Fig. 10) was computed by the following equation:

$$\text{Dopamine selectivity index} = ([\text{DA}]_{\text{left}} - [\text{DA}]_{\text{right}}) / ([\text{DA}]_{\text{left}} + [\text{DA}]_{\text{right}})$$

where $[\text{DA}]_{\text{left}}$ and $[\text{DA}]_{\text{right}}$ represent dopamine concentration during trials to the left and right arms of the maze, respectively.

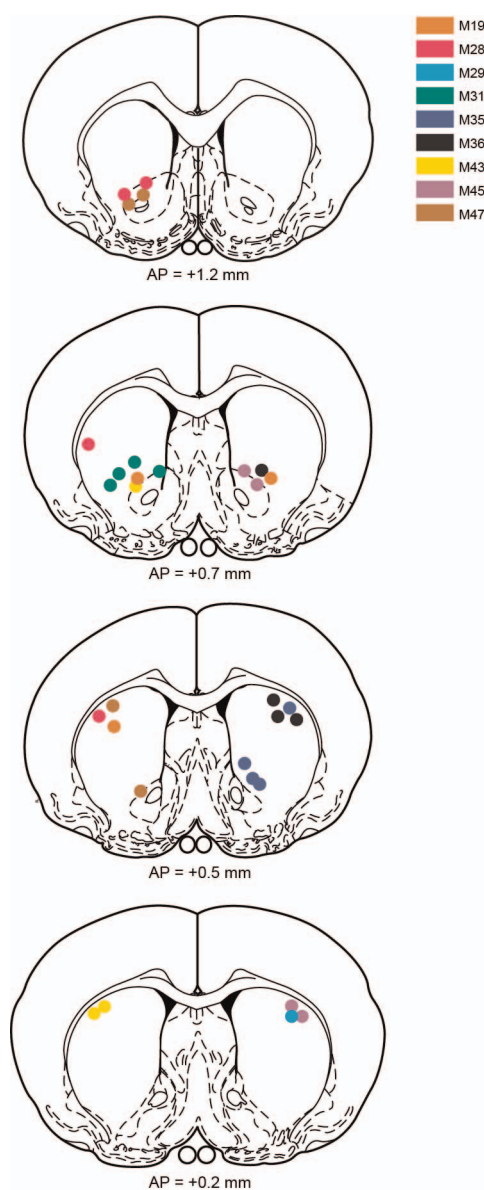
Session-averaged dopamine traces were identified as having positive or negative ramping characteristics (Extended Data Fig. 3) if they exhibited a significantly positive or negative linear regression coefficient (Pearson's, $R > 0.5$ or $R < -0.5$ and $P < 0.01$) over the entire trial period. Trials with phasic responses around the trial start (Extended Data Fig. 4) were identified by calculating the relative difference between consecutive time points (100 ms per sample) in a 1-s window with its centre sliding in 0.1-s steps for a 1-s period from 0.5 s to 1.5 s after warning click. For a given window position, if the differences were all positive values across the first half of the window (0.5 s) and negative across the second half, we determined that a significant inflection point was present in that trial. Comparison of dopamine signals on long and short trials (Fig. 2) was done by selecting trials that fell in the bottom third (short trials) and in the top third (long trials) of the trial-time distribution for each rat that displayed a dopamine peak within 0.5 s of goal-reaching. Trials with noisy video tracking data were discarded from this analysis. The simulations for the time-elapsed model (Fig. 2b, f) were made by calculating the average slope of the ramping signals across all trials on each session and by using linear extrapolation to predict the peak dopamine values on each trial within that session. Peak values were normalized to the median peak value for each session individually and averaged for short and long trials to generate the predictions in Fig. 2f. The multi-transient model (Extended Data Fig. 6) was implemented in MATLAB and tested using a range of physiologically realistic estimates for the slope and decay times for previously observed transient signals *in vivo*¹¹. For the model results shown in Extended Data Fig. 6a, c, simulations were run 100 times, and temporal offset times for the transients were determined by drawing randomly (normrnd function in MATLAB) from a normal distribution with means of 0.8 s (with standard deviation of 0.5 s) and 1.4 s (with standard deviation of 7 s) for short and long trials, respectively (Extended Data Fig. 6a, c). Spatial proximity to goal (Fig. 2c and Extended Data Fig. 7b, e) was calculated by summing the pixel differences in the x and y spatial dimensions for each recorded rat's position. These traces were averaged across all short and long trials separately to generate the traces shown in Fig. 2c. Session-by-session estimations of peak dopamine concentration were made by randomly generating peak trial values using the mean and standard deviation of peak values present in the experimental data. All peak values for short and long trials (Fig. 2f and Extended Data Fig. 6b, d) were normalized to the mean dopamine concentration for each session for both simulated and experimental data.

The presence of population selectivity in the dopamine signals on the associative T-maze task (Fig. 4) was determined by comparing the variance of selectivity indices from the T-maze recordings to the distribution of variances obtained by shuffling the dopamine concentrations on the two end-arms and bootstrapping 10,000 times (Z -test comparing variance of the data to the variances of the bootstrapped distributions, $P < 0.00001$). To identify changes in selectivity and ramping magnitude across training, Z -scores of selectivity indices (Fig. 4f) were computed for each rat by taking the absolute values of all selectivity indices, then by normalizing across sessions before combining all values across rats. Choice selectivity (Extended Data Fig. 10) was computed similarly to the selectivity score for dopamine:

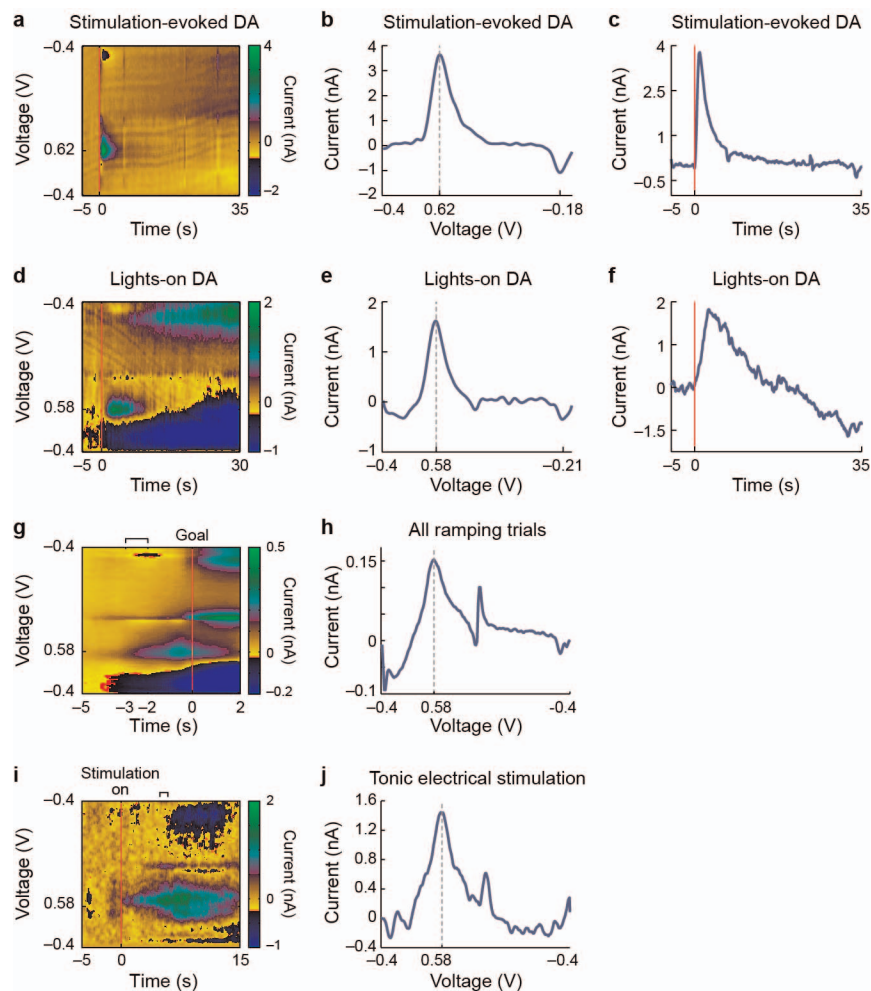
$$\text{Behavioural selectivity index} = (\text{left arm choices} - \text{right arm choices}) / (\text{left arm choices} + \text{right arm choices})$$

Run time biases and per cent correct biases (Extended Data Fig. 10) between the two arms were also computed in this way. Correlations between peak dopamine magnitude and per cent correct performance were calculated by normalizing the average peak dopamine values on each trial to the average peak value across all trials within that session.

Histology. Probe positions were verified histologically¹⁹. Brains were fixed by transcardial perfusion with 4% paraformaldehyde in 0.1 M NaKPO₄ buffer, post-fixed, washed in the buffer solution, and cut transversely at 30 μ m on a freezing microtome, and they were stained with cresylecht violet to allow reconstruction of the recording sites (Extended Data Fig. 1). For a subset of the probes, a constant current (20 mA, 20 s) was passed through the probe before fixation to make micro-lesions at probe-tip locations.

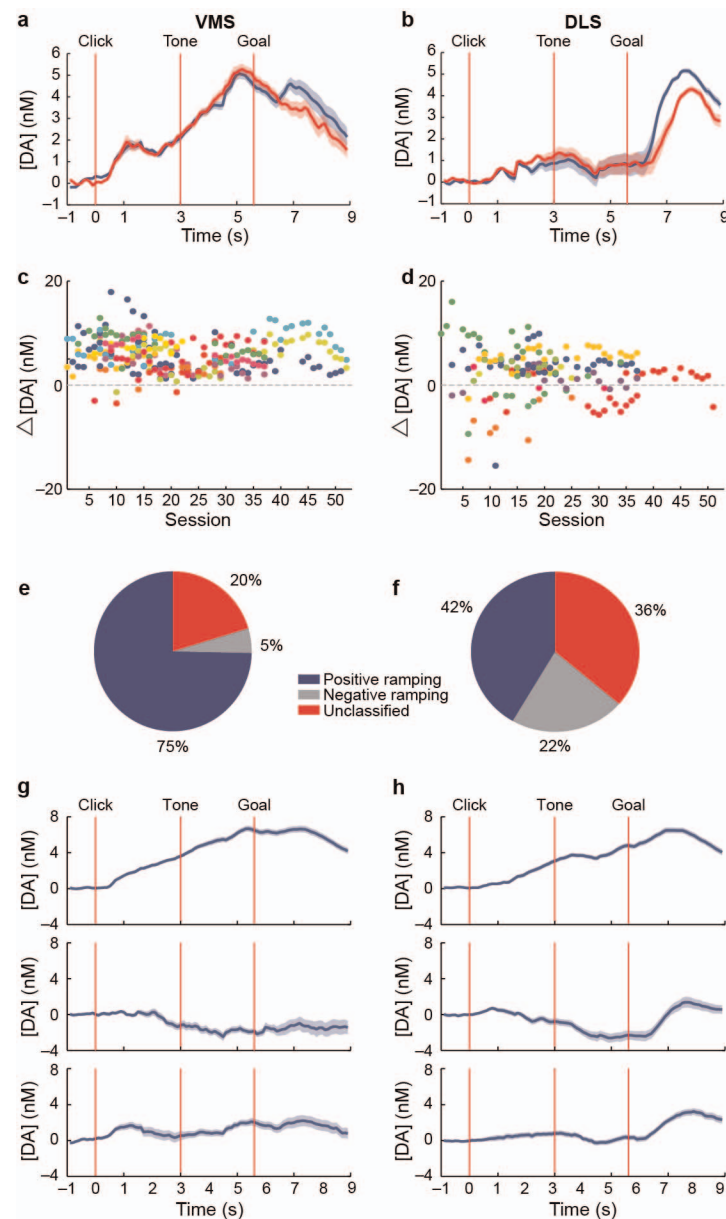


Extended Data Figure 1 | Schematic drawing of recording locations.
Coloured dots indicate positions of probe tips as determined by post-experiment probe length measurements and histology (see Methods). Each colour represents probes from one rat.



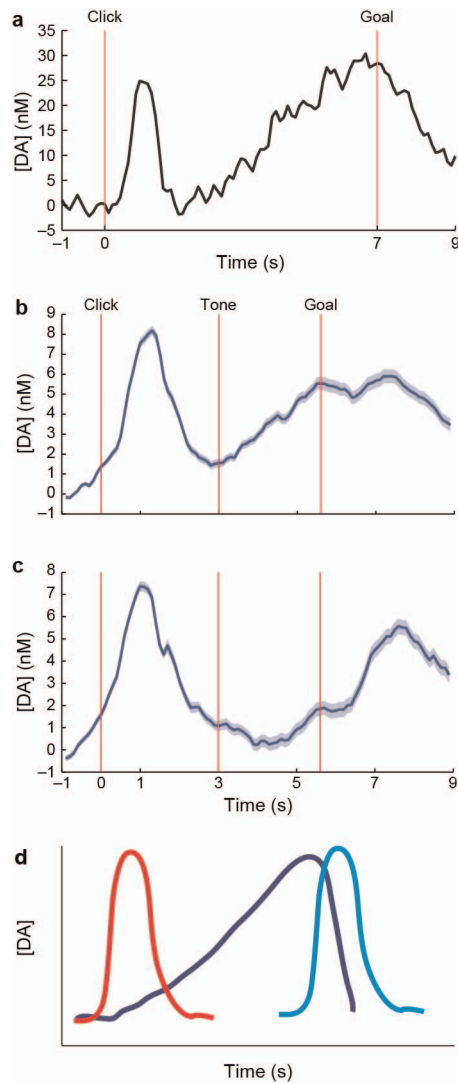
Extended Data Figure 2 | Comparisons of dopamine release in the VMS measured in behaving rats and evoked by electrical MFB stimulation in anaesthetized rats. **a**, Current measured at each potential during successive scans with electrical MFB stimulation (red line) in an anaesthetized rat. **b**, Current–voltage plot from the 5 scans during the 0.5-s period after MFB stimulation. The large increase in current around 0.6 V (dotted line, peak) corresponds to the dopamine redox potential measured *in vitro*. **c**, Time course of current at the dopamine redox potential around electrical MFB stimulation

(red line). **d–f**, Plots, as in **a–c**, illustrating dopamine response in the same rat, now awake, to room lights being turned on. **g**, Current changes averaged over all trials ($n = 4,418$ trials) in which ramping occurred (see Methods) during T-maze running. **h**, Average current–voltage plot for all identified ramping trials for the time period (–3 to –2 s) indicated by the brackets in the colour plot in **g**. **i**, Average current changes induced by tonic MFB stimulation in anaesthetized rats ($n = 3$). **j**, Average current–voltage plot from the bracketed time range (10 to 11 s) in **i** following the onset of stimulation.

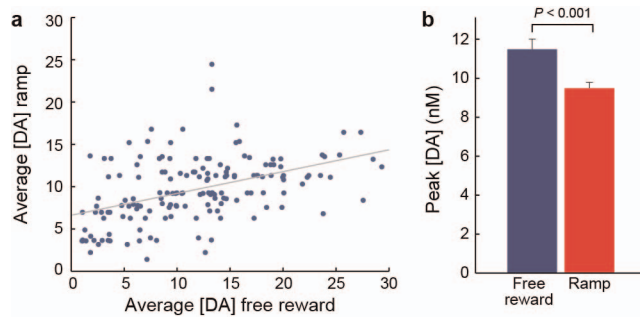


Extended Data Figure 3 | Patterns of dopamine signals in the VMS and DLS. **a, b**, Dopamine concentration recorded in the VMS (**a**, $n = 9$ rats) and in the DLS (**b**, $n = 8$ rats). Data were first averaged across trials to yield session average traces for each probe in each session. These traces were averaged within rat to obtain one average trace per rat, which were then averaged across rats. These plots differ from those in Fig. 1e, f, which considered session averages for each probe to be an independent measure. Shading represents s.e.m. across rats. **c, d**, Distribution of average peak dopamine values for all recordings in VMS

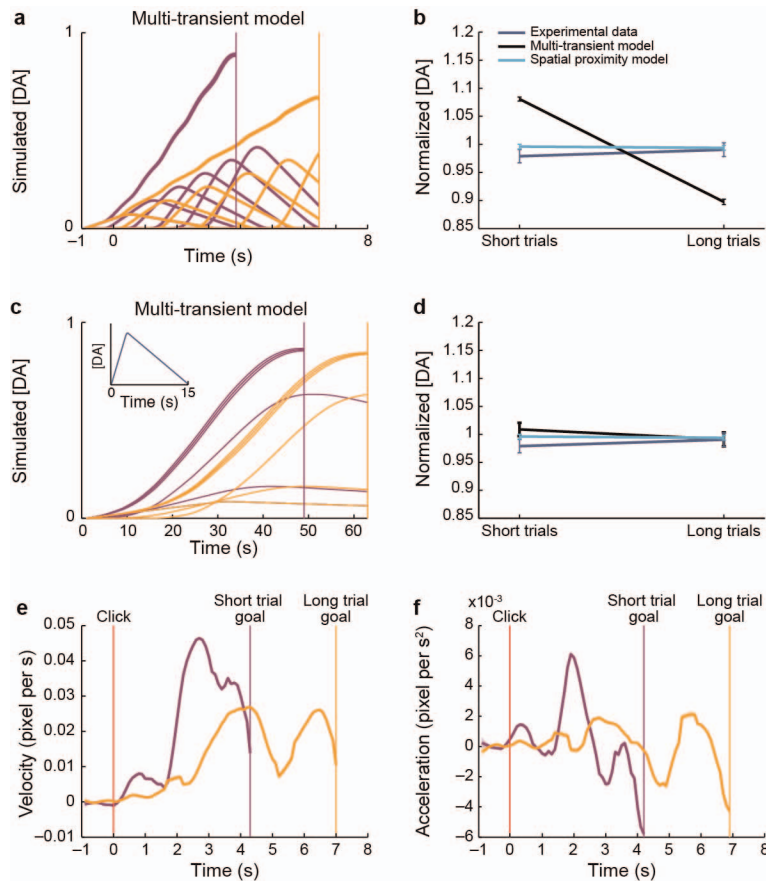
(**c**) and DLS (**d**). Each colour corresponds to an average peak dopamine concentration measured by a single probe in different sessions. **e, f**, Proportion of trial averaged dopamine recordings in the VMS (**e**, out of 300) and DLS (**f**, out of 262) that displayed a positive (blue) or negative (grey) ramping response during maze running, and an unclassified dopamine profile (red). **g, h**, Average dopamine concentration in the VMS (**g**) and DLS (**h**) for the positive ramping traces (top), the negative ramping traces (middle), and the unclassified traces (bottom). Shading represents s.e.m.



Extended Data Figure 4 | Transient responses to cues and rewards occur during the T-maze task performance and are distinct from the sustained ramping response. **a**, Dopamine concentration in a representative trial that included both a phasic response to warning click and a sustained ramping response to goal-reaching. **b**, **c**, Average normalized dopamine traces from VMS (**b**) and DLS (**c**) probes from all trials ($n = 890$ and 640 , respectively) that showed identified transients after warning click (see Methods). In the VMS, note the sharp increase in dopamine around warning click superimposed on the ramping response that followed the phasic click response. Shading represents s.e.m. calculated across trials. **d**, Model for dopamine release profiles in the T-maze task. Sharper transient responses are present at the start of maze running (red) and after goal-reaching (cyan). These responses can be superimposed on and modulated independently of the slower ramping signal related to goal proximity (dark blue).

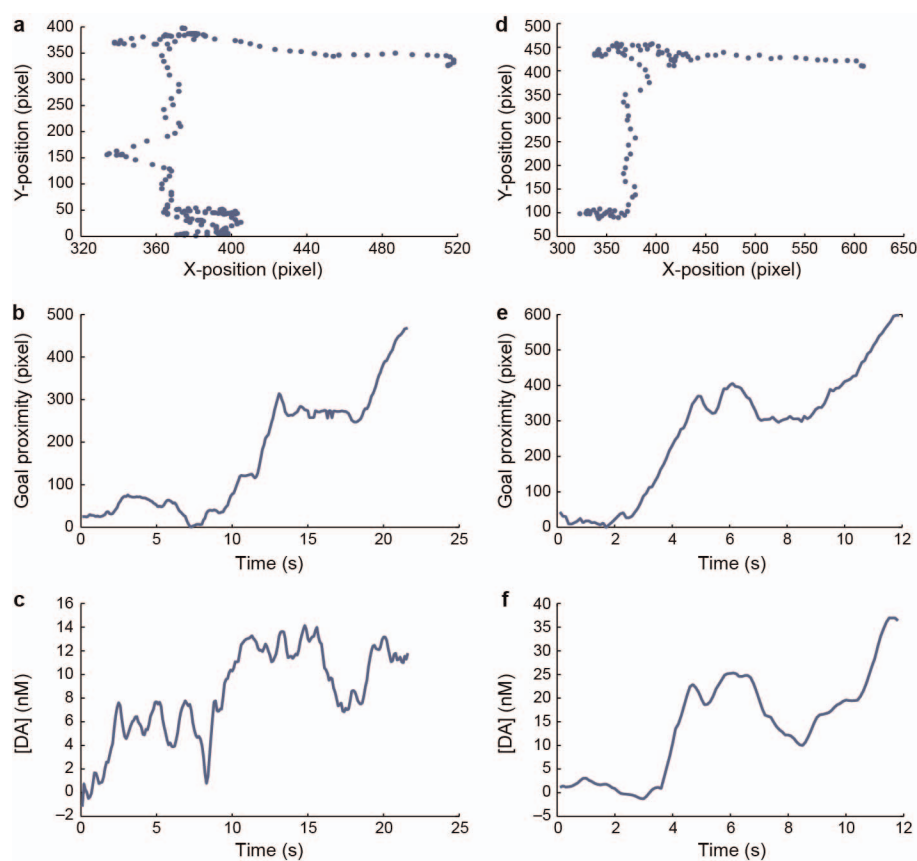


Extended Data Figure 5 | Dopamine signals in response to free reward outside of the task. **a**, Average peak dopamine to unexpected chocolate milk delivery outside the task is positively correlated with peak ramping dopamine measured from the same probes during preceding behavioural training in the maze ($n = 146$ sessions; Pearson's $R = 0.45$, $P < 0.0001$). **b**, Average peak dopamine concentration induced by unexpected free reward outside the task (blue) and peak amplitude of dopamine ramping during maze performance just before reward (red; paired t -test, $P < 0.001$).



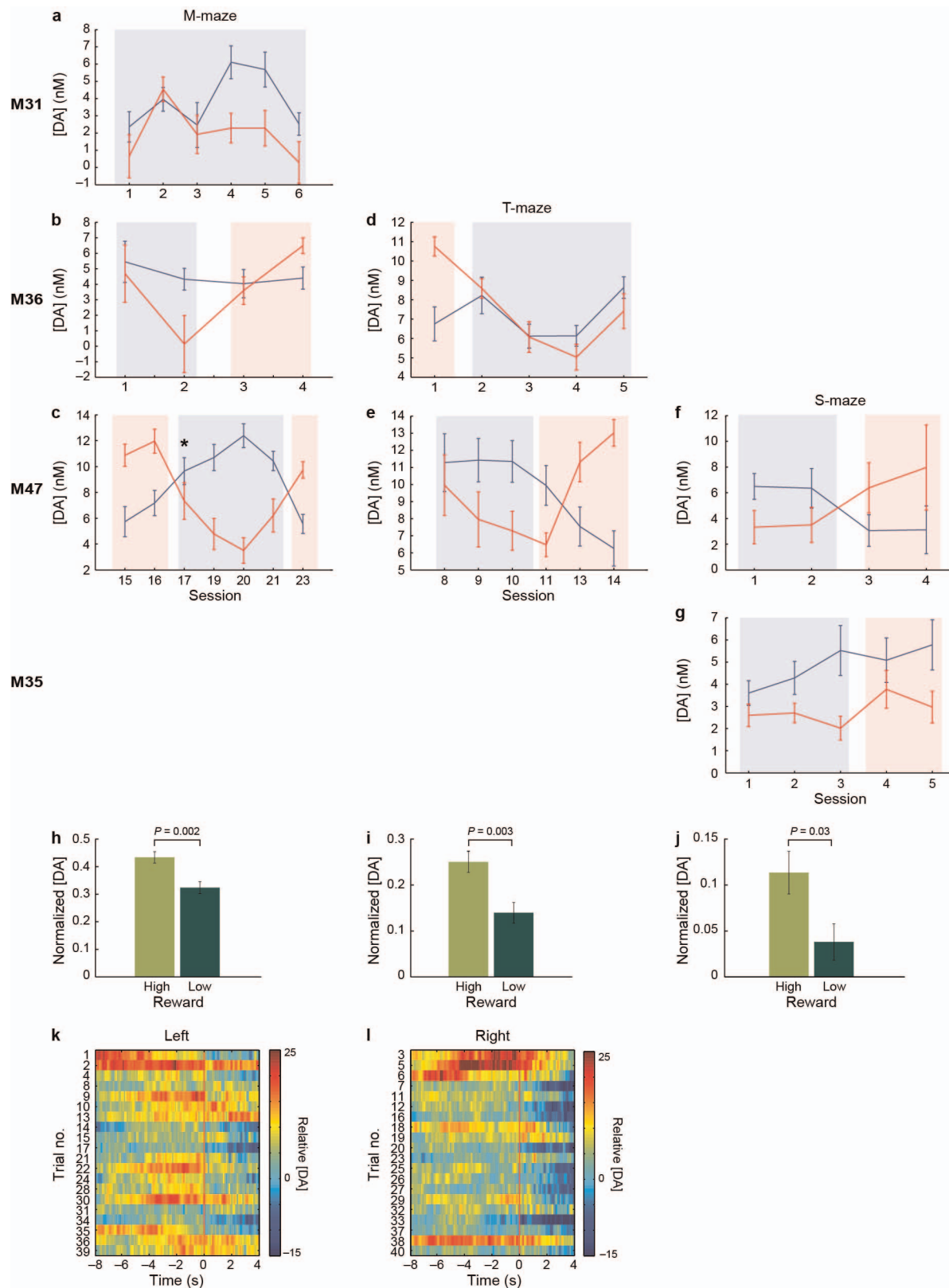
Extended Data Figure 6 | Peak ramping dopamine values do not vary as a function of trial time. **a**, Dopamine release modelled as a summation of four weighted transients in response to fixed maze events on short trials (purple) and long trials (orange). The overlap of the transients is reduced on the long trials, resulting in a lower peak dopamine level at the end of the maze run. Thick lines indicate overall average dopamine, and thin lines indicate the averages of each of the 4 transients across 100 simulation runs. **b**, Relative predicted peak dopamine levels on short and long trials calculated as a linear decay function of trial duration for the simulated model shown in **a** (black), for the spatial proximity model (light blue), and for the actual experimental data (dark blue).

c, An alternative multi-transient model in which the transients (3, inset) are heavily weighted towards the goal location, are highly variable in their time of occurrence, and display a long decay time-course. In this model, the difference between short and long trials is within the noise range of the data. The average of individual transients (inset) across multiple simulations is a smeared version of the single transients that is weighted towards the goal location. **d**, Data plotted as in **b**, for the alternative multi-transient model. **e**, **f**, Average run speed (**e**) and acceleration (**f**) during short (purple) and long (orange) trials, as shown in Fig. 2, for all animals and sessions.



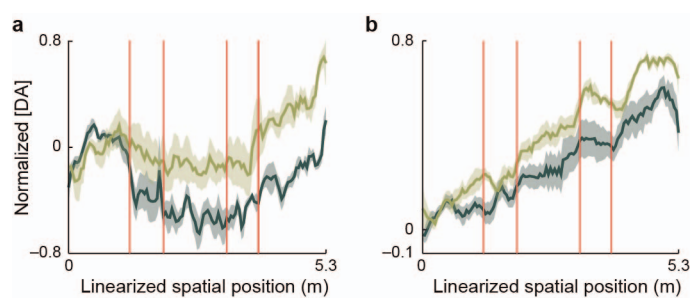
Extended Data Figure 7 | Examples of trials with pausing or exploration near the choice point of the T-maze. **a–c,** Video tracker traces (**a**), relative proximity to reward (**b**), and dopamine concentration (**c**) measured during a

single trial in which a rat paused near the choice point. **d–f,** Video tracker traces (**d**), relative proximity to reward (**e**) and dopamine concentration (**f**) measured during another trial from a different rat.

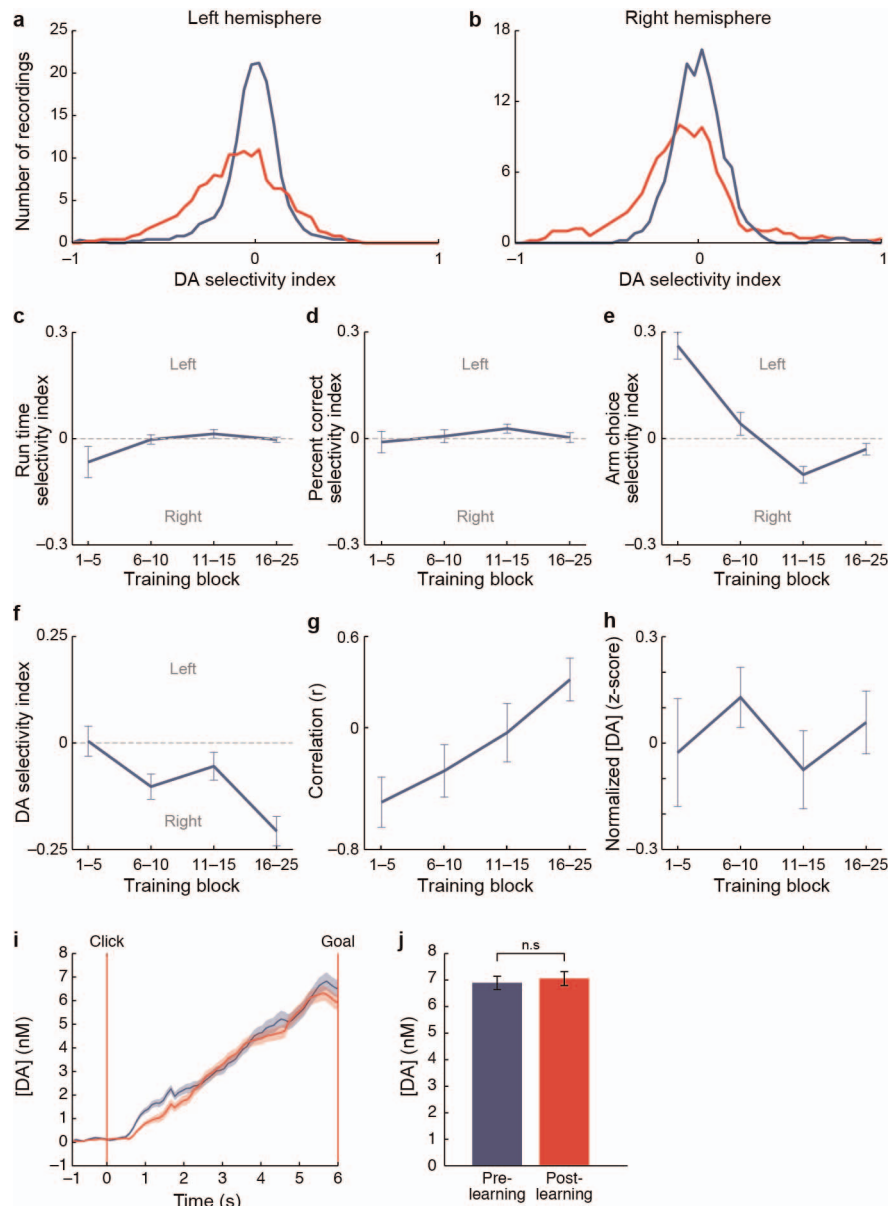


Extended Data Figure 8 | Value manipulation experiments in the T-maze, M-maze, and S-maze for all rats. **a–c**, Average peak dopamine levels for M-maze sessions for three individual rats (M31, M36, and M47) in the left (blue) and right (red) end-arms. Blue and red shading indicates sessions in which left and right arm contained the larger reward, respectively. Error bars indicate s.e.m. **d, e**, Average peak dopamine levels, as in **a–c**, for value bias T-maze sessions for two rats (M36 and M47). **f, g**, Average peak dopamine

levels, as in **a–c**, for value bias S-maze sessions for 2 rats (M47 and M35). **h–j**, Average normalized dopamine levels measured in the high reward arm (light green) and low reward arm (dark green) as rats performed the M-maze (**h**), T-maze (**i**) and S-maze (**j**) tasks. **k, l**, Dopamine concentration relative to left (**k**) and right (**l**) goal-reaching during the first session following a reversal of reward values (session 17 of M47, indicated by an asterisk in **c**).



Extended Data Figure 9 | Dopamine signals recorded in two rats trained on the S-maze task. **a**, Average normalized dopamine concentrations measured from VMS probes in one rat ($n = 5$ sessions) performing the S-maze task as in Fig. 3j. Light green line indicates runs to the higher reward goal, and dark green lines to the lower reward goal. Shading indicates s.e.m. Red vertical lines indicate turns. **b**, Traces, as in **a**, for the second rat trained on the S-maze task ($n = 4$ sessions).



Extended Data Figure 10 | Dopamine end-arm preferences become positively correlated with arm choices with training, but ramping dopamine signals do not change significantly with performance improvement. **a**, **b**, Distribution of selectivity indices, as in Fig. 4d, for all probes implanted in the left (**a**, $n = 5$ rats) and right (**b**, $n = 3$ rats) hemispheres. Note the bias in both groups of the selectivity preference towards negative selectivity indices (right bias, red) relative to the shuffled data (blue). **c–e**, Biases in average run time (**c**), percentage of correct responses (**d**), and arm choices (**e**) across training blocks. Negative values indicate biases towards the right end-arm. **f**, Raw average dopamine selectivity indices across training blocks. Note emergence of right bias with training. **g**, Correlation coefficients (Pearson's R) computed for each training block between arm choice selectivity indices and dopamine

selectivity indices. Error bars indicate confidence limits of the correlations. **h**, Normalized peak magnitudes of dopamine signals averaged in a 0.5-s window before goal-reaching in sessions with significant (Mann–Whitney U -test, $P < 0.05$) pre-goal increases. Data are averaged across rats for each training block (for which, left to right, $n = 48, 101, 113$ and 179 trial averaged recordings, respectively). **i**, Average (\pm s.e.m.) dopamine concentration from ramping dopamine sessions in which percentage of correct trials fell above (red, $n = 179$) or below (blue, $n = 92$) the learning criterion for T-maze task acquisition (72.5% correct, chi-square test, $P < 0.05$). **j**, Average (\pm s.e.m.) peak dopamine levels from the sessions plotted in **i**, showing no significant difference between pre- and post-learning periods (t -test, $P = 0.44$).

A gustatory receptor paralogue controls rapid warmth avoidance in *Drosophila*

Lina Ni^{1,2}, Peter Bronk^{1,2}, Elaine C. Chang^{1,2}, April M. Lowell^{1,2}, Juliette O. Flam^{1,2}, Vincent C. Panzano^{1,2}, Douglas L. Theobald³, Leslie C. Griffith^{1,2} & Paul A. Garrity^{1,2}

Behavioural responses to temperature are critical for survival, and animals from insects to humans show strong preferences for specific temperatures^{1,2}. Preferred temperature selection promotes avoidance of adverse thermal environments in the short term and maintenance of optimal body temperatures over the long term^{1,2}, but its molecular and cellular basis is largely unknown. Recent studies have generated conflicting views of thermal preference in *Drosophila*, attributing importance to either internal³ or peripheral⁴ warmth sensors. Here we reconcile these views by showing that thermal preference is not a singular response, but involves multiple systems relevant in different contexts. We found previously that the transient receptor potential channel TRPA1 acts internally to control the slowly developing preference response of flies exposed to a shallow thermal gradient³. We now find that the rapid response of flies exposed to a steep warmth gradient does not require TRPA1; rather, the gustatory receptor GR28B(D) drives this behaviour through peripheral thermosensors. Gustatory receptors are a large gene family, widely studied in insect gustation and olfaction, and are implicated in host-seeking by insect disease vectors^{5–7}, but have not previously been implicated in thermosensation. At the molecular level, GR28B(D) misexpression confers thermosensitivity upon diverse cell types, suggesting that it is a warmth sensor. These data reveal a new type of thermosensory molecule and uncover a functional distinction between peripheral and internal warmth sensors in this tiny ectotherm reminiscent of thermoregulatory systems in larger, endothermic animals². The use of multiple, distinct molecules to respond to a given temperature, as observed here, may facilitate independent tuning of an animal's distinct thermosensory responses.

Thermal preference is an important body temperature control mechanism from insects to humans^{1,2}. In *Drosophila* two sets of warmth-sensing neurons (activated above ~25 °C) have been proposed to control thermal preference: the anterior cell (AC) neurons³, located inside the head, and the hot cell (HC) neurons⁴, located peripherally in the arista (Fig. 1a). However, different studies suggest conflicting cellular and molecular mechanisms for thermal preference control. At the cellular level, primary importance has been attributed to either internal³ or peripheral⁴ warmth sensors. At the molecular level, the internal AC neurons sense warmth via *TrpA1* (ref. 3), which encodes a warmth-activated transient receptor potential (TRP) channel^{3,8}, whereas the peripheral HC neurons seem to be *TrpA1*-independent⁴. To clarify the mechanisms of thermal preference, we sought to discover the molecular basis of HC neuron function.

The arista contains six neurons⁹: three warmth-responsive HC neurons (which can be labelled using cell-specific Gal4 expression in the *HC-GAL4* strain⁴) and three cool-responsive (cold cell; CC) neurons (labelled in the *CC-GAL4* strain⁴) (Fig. 1b–d). Three unidentified cells in the arista have been reported to express *Gr28b.d-GAL4*, a transgene in which promoter sequences upstream of the gustatory receptor GR28B(D) control Gal4 expression¹⁰. We found that these *Gr28b.d-GAL4*-expressing cells resembled thermoreceptors, with cell bodies

near the arista base and thin processes in the shaft (Fig. 1e). To determine the thermoreceptor subset labelled, *Gr28b.d-GAL4* was combined with each thermoreceptor-specific Gal4. *Gr28b.d-GAL4* plus *HC-GAL4* labelled three neurons (Fig. 1f, *n* = 5), whereas *Gr28b.d-GAL4* plus *CC-GAL4* labelled six neurons (Fig. 1g, *n* = 5), indicating that *Gr28b.d-GAL4* is expressed in the HC neurons. Although *in situ* hybridization was unsuccessful (common for gustatory receptors⁵), GR28B(D) transcripts were robustly detected in dissected antennae/aristae from wild-type, but not *Gr28b* mutant, animals by reverse transcriptase PCR (RT-PCR) (Supplementary Fig. 1), demonstrating expression in this tissue.

Gustatory receptors are a large family of seven transmembrane proteins present in invertebrates⁷, with 68 members in *Drosophila melanogaster*¹¹ (Supplementary Fig. 2). Insects also contain multiple gustatory receptor-related odorant receptors (62 in *D. melanogaster*¹¹). Gustatory receptors and odorant receptors form a gene family distinct from, and apparently unrelated to, the G-protein-coupled receptor superfamily⁷. Gustatory receptors and odorant receptors have been studied extensively as chemoreceptors for sweet and bitter tastants, food odours, carbon dioxide and other chemicals^{5–7}, but have not previously been implicated in thermosensation. We examined gustatory receptor involvement in thermosensation using a two-temperature choice assay¹², exposing flies for 1 min to a steep thermal gradient (initially >5 °C per cm) created using tubes of ~25.5 and ~31.0 °C air (a preferred and an increased-but-innocuous temperature, respectively) separated by 1 cm. Flies normally prefer the cooler tube, a behaviour termed 'rapid negative thermotaxis' (Fig. 1h, i). Consistent with a previous report⁴, inhibiting HC neurons by cell-specific expression of tetanus toxin light chain (TNT), a vesicle release inhibitor¹³, using *HC-GAL4* strongly reduced such behaviour (Fig. 1h). In agreement with the importance of HC neurons, and in addition to previous studies¹⁴, third antennal segment/arista removal strongly reduced this behaviour, whereas ablating other tissues expressing *HC-GAL4* and *Gr28b.d-GAL4* did not (Supplementary Figs 3–5). By contrast, inhibiting AC neurons by TNT expression using *TrpA1^{GAL4}*, a Gal4 knock-in at the *TrpA1* locus¹⁵, had no effect (Fig. 1h). (This manipulation disrupted a previously reported AC-dependent thermosensory behaviour³ (Supplementary Fig. 6).) These data indicate that rapid negative thermotaxis depends on the peripheral HC warmth sensors.

To probe the molecular basis of rapid negative thermotaxis, we first examined its dependence on *TrpA1*, which is required for AC neuron warmth-sensing³. Consistent with the *TrpA1*-independence of HC neuron thermosensitivity⁴, a strong loss-of-function *TrpA1* mutation did not affect this behaviour (Fig. 1i). By contrast, strong loss-of-function mutations in the gene encoding GR28B(D) eliminated the response; *Gr28b* mutants distributed nearly equally between ~25.5 °C and ~31.0 °C (Fig. 1k). The defect was specific: excising the transposon in the *Gr28b^{Mi}* allele restored thermotaxis (Fig. 1k), and both a *Gr28b*-containing genomic transgene and *Gr28b(D)* complementary DNA expression rescued the mutant (Fig. 1k, l). We also attempted rescue

¹National Center for Behavioral Genomics, Brandeis University, Waltham, Massachusetts 02454, USA. ²Volen Center for Complex Systems, Department of Biology, Brandeis University, Waltham, Massachusetts 02454, USA. ³Department of Biochemistry, Brandeis University, Waltham, Massachusetts 02454, USA.

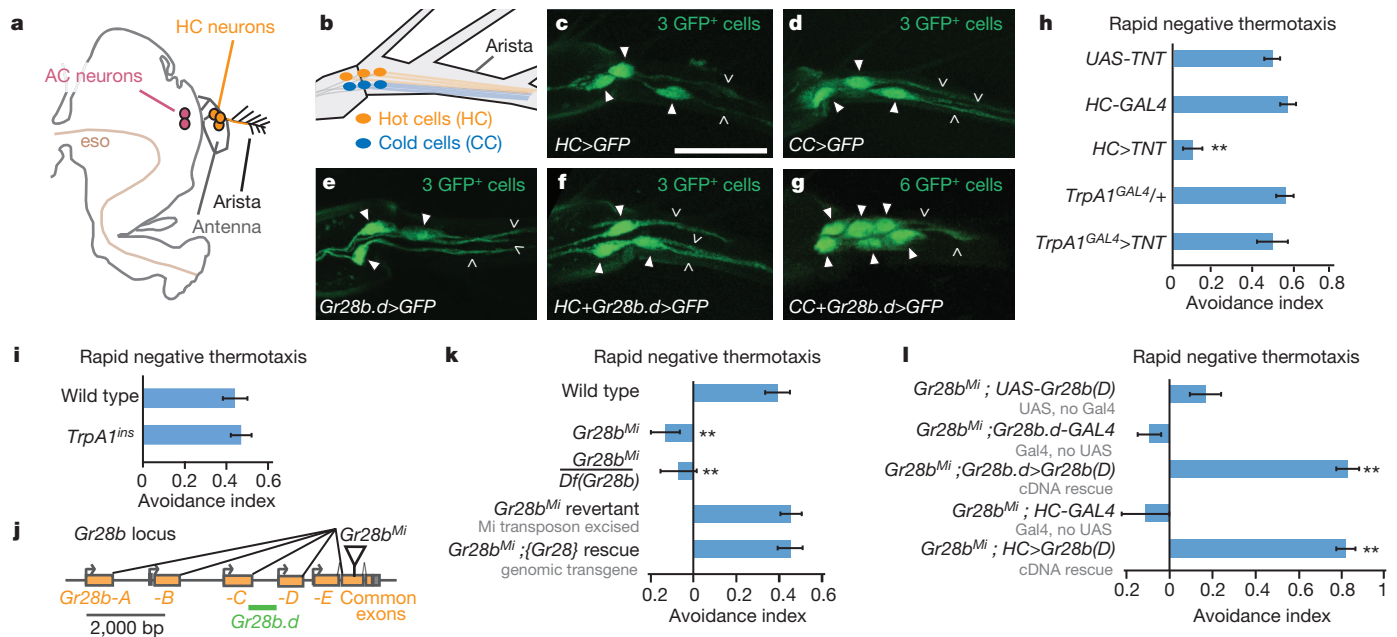


Figure 1 | *Gr28b* is required for rapid negative thermotaxis. **a**, Head. eso, oesophagus. **b**, Arista. **c–g**, *HC-GAL4;UAS-nls:GFP* (in which *UAS* denotes upstream activation sequence and *nls* denotes nuclear localization signal) (**c**), *CC-GAL4;UAS-nls:GFP* (**d**), *Gr28b.d-GAL4;UAS-nls:GFP* (**e**), *HC-GAL4;Gr28b.d-GAL4;UAS-nls:GFP* (**f**), *CC-GAL4;Gr28b.d-GAL4;UAS-nls:GFP* (**g**). Arrowheads, cell bodies; carats, processes. Scale bar, 15 μ m. **h**, *UAS-TNT* ($n = 22$), *HC-GAL4* ($n = 15$), *HC>TNT* ($n = 29$), *TrpA1^{GAL4/+}* ($n = 9$), *TrpA1^{GAL4>TNT}* ($n = 9$). **i**, Wild type ($n = 12$), *TrpA1^{ins}* ($n = 29$). **j**, *Gr28b*. Arrows denote isoform-specific start sites. Green line, *Gr28b.d-GAL4*

by expressing cDNAs for the other *Drosophila* GR28 family members^{10,11} (four other *Gr28b* isoforms (Fig. 1j) and *Gr28a*^{11,12}) under *Gr28b.d-GAL4* control. Although a negative result could reflect a failure to be properly expressed, only *Gr28b(E)* yielded significant rescue (Supplementary Fig. 7). However, endogenous *Gr28b(E)* transcripts were not detected in the antenna/arista (Supplementary Fig. 1), consistent with a previous analysis indicating that GR28B(E) is not expressed there¹⁰. Together, these data demonstrate that rapid negative thermotaxis depends not on *TrpA1*, but on *Gr28b*, consonant with the specific dependence of this behaviour on HC neuron function (Fig. 1h). Notably, cell-specific GR28B(D) expression using *HC-GAL4* strongly rescued the *Gr28b* mutant (Fig. 1l), indicating that GR28B(D) function in the HC thermosensors is sufficient to restore rapid negative thermotaxis.

To test whether GR28B(D) might act as a thermosensor, we examined whether it conferred warmth-sensitivity when ectopically expressed. Unlike controls, flies broadly expressing GR28B(D) under *Actin5C-GAL4* control were incapacitated when heated to 37 °C for 3 min, recovering when returned to 23 °C (Fig. 2a and Supplementary Video 1). This dramatic effect suggested that GR28B(D) might promote warmth-responsive neuronal activation. We showed previously that ectopic expression of the warmth-activated cation channel TRPA1(B), a product of *Drosophila* *TrpA1*, renders fly chemosensors warmth-responsive¹⁶. Like TRPA1(B), chemosensor expression of GR28B(D) (using *Gr5a-GAL4*) conferred robust warmth-responsiveness (Fig. 2b). We examined the behavioural consequences of such GR28B(D) expression. When chemically activated, sweet-responsive chemosensors promote proboscis extension^{5,6}. When GR28B(D) was expressed in these cells, strong proboscis extension was elicited by warming to ~32 °C (Fig. 2c). This ability to confer warmth-responsiveness is consistent with GR28B(D) acting as a warmth sensor.

Whether GR28B(D) requires sensory neuron-specific cofactors was examined in the neuromuscular system. Unlike controls, motor neurons expressing GR28B(D) (using *OK371-GAL4*) triggered warmth-responsive excitatory junction potentials at the neuromuscular junction

(Fig. 3a). Thus, GR28B(D)-mediated warmth-responsiveness does not require sensory neuron-specific cofactors. The threshold for GR28B(D)-dependent muscle stimulation was 26.0 ± 0.3 °C (\pm s.e.m., $n = 12$), just above TRPA1(B)'s ~25 °C threshold in this system¹⁷, indicating that both molecules mediate responses to innocuous warming.

To quantify the thermosensitivity of GR28B(D)-dependent responses, currents were monitored using whole-cell patch-clamp electrophysiology. Unlike controls, voltage-clamped motor neurons expressing GR28B(D) exhibited warmth-responsive inward currents (Fig. 3b). The response's temperature coefficient (Q10, fold change in current per 10 °C change) was calculated by Arrhenius analysis¹⁸ (Fig. 3c). GR28B(D)-dependent currents were highly thermosensitive (Q10 of 25 ± 5 (s.e.m., $n = 7$)), similar to mammalian neurons expressing thermosensitive TRP channels¹⁸. Substituting *N*-methyl-D-glucamine (NMDG)⁺ for Na⁺ in the extracellular solution eliminated heat-responsiveness, consistent with cation channel activation ($n = 3$; Supplementary Fig. 8).

The potential dependence of GR28B(D) on neuron-specific cofactors was tested in muscle. Although control muscles voltage-clamped at -60 mV exhibited modest warmth-responsive outward currents (Fig. 3d), muscles expressing GR28B(D) (using *Mhc-GAL4*) exhibited robust warmth-responsive inward currents (Fig. 3d). The ability of GR28B(D) to confer warmth sensitivity across diverse cell types supports the hypothesis that GR28B(D) acts as a molecular thermoreceptor. It further suggests GR28B(D) as a new class of tool for thermogenetic neuronal activation, adding to the TRP-based toolbox currently used in *Drosophila*¹⁹.

Although GR28B(D) resembles TRPA1(B) in conferring warmth-sensitivity^{3,16,17}, these two proteins have distinct functions in the fly, with only *Gr28b* controlling rapid negative thermotaxis (Fig. 1). *TrpA1* was found previously to control the slowly developing thermal preference response of flies on a shallow, broad thermal gradient (~0.5 °C per cm, 18–32 °C)³. We tested the contribution of *Gr28b* to this long-term body temperature selection behaviour. As reported previously³, *TrpA1* mutants selected unusually warm temperatures after 30 min on

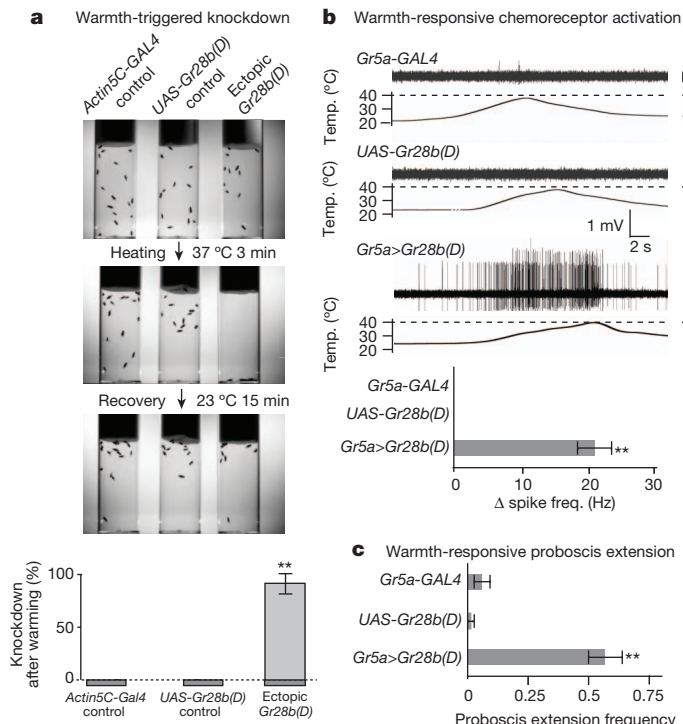


Figure 2 | GR28B(D) expression confers warmth-responsive neuronal activation and behaviour. **a**, Top, flies before and after warming. Bottom, knockdown quantification ($n = 3$ independent assays per genotype, >10 flies per assay). Ectopic *Gr28b(D)* denotes *Actin5C-GAL4;UAS-Gr28b(D)*. **b**, Gustatory bristle responses to warming. Top, extracellular recording traces. Bottom, average spike rate from gustatory bristles during warming, after subtracting electrolyte-only baseline. *Gr5a-GAL4* ($n = 6$) bristles; *UAS-Gr28b(D)* ($n = 9$), *Gr5a>Gr28b(D)* ($n = 17$). **c**, Frequency of proboscis extension upon warming to $\sim 32^\circ\text{C}$ ($n = 32$ flies per genotype). Data are mean \pm s.e.m. **Significantly different from UAS and Gal4 alone controls (Tukey's HSD, $\alpha = 0.01$).

the gradient, with many accumulating at $\geq 28^\circ\text{C}$ (Fig. 4a). By contrast, strong loss-of-function *Gr28b* mutants behaved indistinguishably from wild type (Fig. 4a). This neatly distinguishes *Gr28b* and *TrpA1*, with the former controlling rapid negative thermotaxis and the latter long-term body temperature selection.

These findings reconcile previously disparate views of *Drosophila* thermosensation^{3,4} by demonstrating that thermal preference is not a

singular behaviour, but involves multiple systems relevant in different contexts. It suggests a model in which *Gr28b*, acting peripherally, controls rapid responses to ambient temperature jumps, whereas *TrpA1*, acting internally, controls responses to sustained temperature increases reaching the core. In the arista, *Gr28b* could experience ambient temperature fluctuations in advance of core changes, eliciting rapid avoidance. Such behaviour could be critical for a tiny animal in which ambient and core temperatures equalize rapidly¹. The dispensability of *Gr28b* for responses on the shallow gradient (Fig. 4a) could relate to observations in other insects where peripheral thermoreceptors respond more to temperature fluctuations than absolute values²⁰. The fly's reliance on distinct sensors for distinct aspects of thermal preference is reminiscent of complex thermosensory systems of larger, endothermic animals². In the fly, these warmth-responsive pathways potentially converge in the brain, where both sets of sensors innervate overlapping regions⁴.

Finally, we tested whether *Gr28b* and *TrpA1* were uniquely suited to their roles in the fly. Although *TrpA1* was normally not required for rapid negative thermotaxis (Fig. 1i), when expressed in the arista using *Gr28b.d-GAL4*, TRPA1(B) significantly rescued the *Gr28b* mutant defect (Fig. 4b). (As expected, a less thermosensitive *TrpA1* isoform, TRPA1(A)¹⁶, did not rescue the defect (Fig. 4b).) Conversely, although *Gr28b* was not normally required for slowly developing thermal preference on the shallow gradient (Fig. 4a), GR28B(D) expression under *TrpA1^{GAL4}* control significantly rescued the *TrpA1* mutant defect (Fig. 4c). Thus, when their expression is manipulated appropriately, GR28B(D) and TRPA1(B) can act in the same cells and support the same behaviours, indicating fundamental functional similarities.

Although studied extensively, the mechanisms of gustatory receptor action are not fully resolved⁷. Gustatory receptors have been reported to act as cation channels^{7,21} and via G-proteins²². Whether GR28B(D) acts by either mechanism remains unknown. Although attempts to study GR28B(D) in heterologous cells (including *Xenopus laevis* oocytes and HEK cells; L.N., T. Lauer, P. Taneja, S. Nelson and P.A.G., unpublished observations) were unsuccessful, the ability of GR28B(D) to confer warmth-responsiveness upon diverse cell types argues against a requirement for cell-type-specific cofactors in the fly. *Gr28b* has been implicated in responses to strong illumination²³. This seems to be unrelated to GR28B(D)-dependent thermosensation, as *Gr28b*-dependent photosensors are unresponsive to innocuous warming²³ and appear to express other *Gr28b* isoforms¹⁰. GR28B(D)-expressing muscles were not light-responsive ($n = 4$, Supplementary Fig. 9).

Previous studies have demonstrated the importance of TRP channels in *Drosophila* thermosensation¹, stimulating interest in their

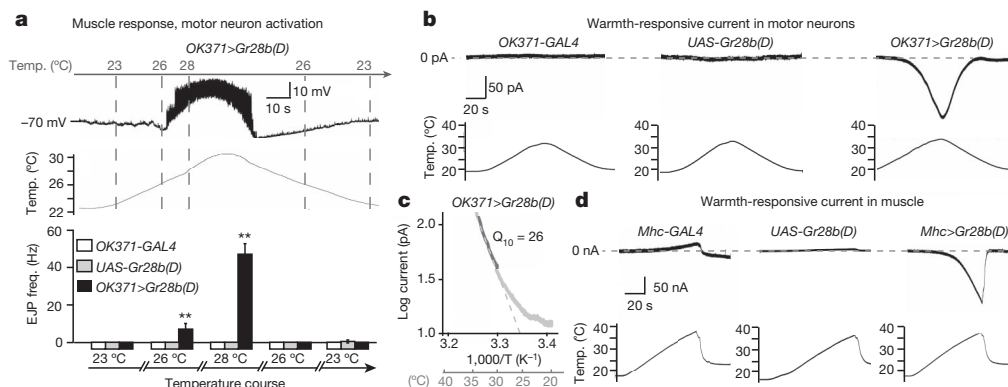


Figure 3 | GR28B(D) expression yields highly thermosensitive currents. **a**, Top, muscle response to warming in *OK371>Gr28b(D)* animals. Bottom, excitatory junction potentials (EJP) during temperature course. *OK371-GAL4* ($n = 12$) muscles; *UAS-Gr28b(D)* ($n = 13$), *OK371>Gr28b(D)* ($n = 9$). **b**, Currents in voltage-clamped motor neurons upon warming. *OK371-GAL4* ($n = 5$) motor neurons; *UAS-Gr28b(D)* ($n = 5$), *OK371>Gr28b(D)* ($n = 7$).

c, Arrhenius plot of warmth-responsive current of *OK371>Gr28b(D)* motor neuron in panel **b**. **d**, Currents in voltage-clamped muscles upon warming. *Mhc-GAL4* ($n = 3$) muscles; *UAS-Gr28b(D)* ($n = 3$), *Mhc>Gr28b(D)* ($n = 7$). Data are mean \pm s.e.m. **Significantly different from UAS and Gal4 alone controls (Tukey's HSD, $\alpha = 0.01$).

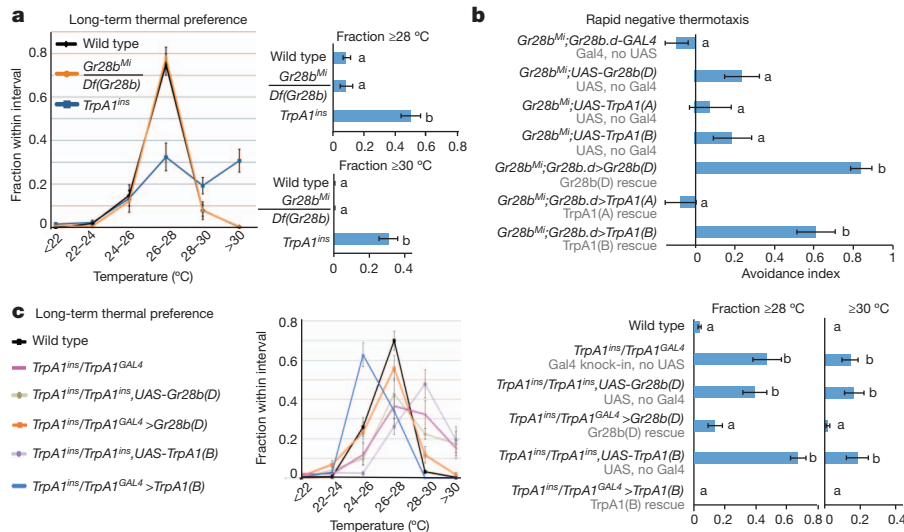


Figure 4 | GR28B(D) and TRPA1(B) functionally substitute for one another. **a**, Left, fly distribution across gradient (30 min). Right, flies in $\geq 28^{\circ}\text{C}$ and $\geq 30^{\circ}\text{C}$ regions. Wild type ($n = 11$), $\text{Gr28b}^{Mi}/\text{Df}(\text{Gr28b})$ ($n = 8$), TrpA1^{ins} ($n = 17$). **b**, Rapid negative thermotaxis. $\text{Gr28b}^{Mi};\text{Gr28b.d-GAL4}$ ($n = 14$), $\text{Gr28b}^{Mi};\text{UAS-Gr28b(D)}$ ($n = 10$), $\text{Gr28b}^{Mi};\text{UAS-TrpA1(A)}$ ($n = 8$), $\text{Gr28b}^{Mi};\text{UAS-TrpA1(B)}$ ($n = 9$), $\text{Gr28b}^{Mi};\text{Gr28b.d>Gr28b(D)}$ ($n = 7$), $\text{Gr28b}^{Mi};\text{Gr28b.d>TrpA1(A)}$ ($n = 9$), $\text{Gr28b}^{Mi};\text{Gr28b.d>TrpA1(B)}$ ($n = 9$)

(GAL4 alone, UAS-Gr28b(D) alone and Gr28b(D) rescue data from Fig. 11.) **c**, Data presented as in **a**. Wild type ($n = 11$), $\text{TrpA1}^{ins}/\text{TrpA1}^{GAL4}$ ($n = 12$), $\text{TrpA1}^{ins};\text{UAS-Gr28b(D)}$ ($n = 11$), $\text{TrpA1}^{ins}/\text{TrpA1}^{GAL4}>\text{Gr28b(D)}$ ($n = 14$), $\text{TrpA1}^{ins};\text{UAS-TrpA1(B)}$ ($n = 6$), $\text{TrpA1}^{ins}/\text{TrpA1}^{GAL4}>\text{TrpA1(B)}$ ($n = 5$). Data are mean \pm s.e.m.; n , number of independent assays. Letters denote statistically distinct groups (Tukey's HSD, $\alpha = 0.01$).

potential involvement in warmth-dependent host-seeking by insect disease vectors²⁴. This work raises the possibility that gustatory receptors, including GR28 receptors in disease vectors such as tsetse flies and mosquitoes (Supplementary Fig. 2), regulate thermosensation more broadly. GR28B(D) adds to a growing list of highly thermosensitive membrane proteins including not only TRPs, but the mammalian ANO1 chloride channel²⁵ and calcium-channel regulator STIM1 (ref. 26). The presence of exceptional thermosensitivity in diverse proteins may facilitate temperature-responsive modulation of diverse physiological responses. Furthermore, using multiple molecules to mediate behavioural responses to similar temperatures may facilitate independent tuning of distinct thermosensory responses.

METHODS SUMMARY

Fly strains. Gr28b , TrpA1 , HC-GAL4 and CC-GAL4 strains were previously described^{3,4,10,15,16,23,27}. $\text{Df}(\text{Gr28b})$ is $\text{Df}(2\text{L})\text{Exel7031}$ (Bloomington Stock Center). To control for transposon position effects, all UAS-Gr28 transgenes were inserted at the same landing site, attP2, by site-specific transgenesis¹⁶.

Behaviour and physiology. Two-temperature rapid negative thermotaxis assay was as described¹², except tube temperatures were $25.5 \pm 0.3^{\circ}\text{C}$ and $31.0 \pm 0.5^{\circ}\text{C}$ (\pm s.d.), ≥ 15 flies per trial. Thermal preference assay was as described^{3,12}, with 20–60 flies (2–5 days old) per trial. For proboscis extension, female flies (1–5 days old) were starved overnight with water, then glued to glass slides and heated¹⁶. Flies were given three 5-s heat presentations at 5-s intervals. Physiology is detailed in methods.

Molecular biology. Transgenic fly creation and RT-PCR were performed as described¹⁶. RT-PCR primers straddled splice junctions to minimize genomic DNA amplification. Three independent tissue preparations gave similar results.

Phylogeny. As gustatory receptor sequence diversity creates the potential for alignment ambiguities, phylogeny was created using BALi-Phy²⁸, which performs simultaneous Bayesian inference of alignment and phylogeny. Further details provided in Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 7 January; accepted 14 June 2013.

Published online 7 August 2013.

- Garrity, P. A., Goodman, M. B., Samuel, A. D. & Sengupta, P. Running hot and cold: behavioral strategies, neural circuits, and the molecular machinery for thermotaxis in *C. elegans* and *Drosophila*. *Genes Dev.* **24**, 2365–2382 (2010).
- Flouris, A. D. Functional architecture of behavioural thermoregulation. *Eur. J. Appl. Physiol.* **111**, 1–8 (2011).

- Hamada, F. N. et al. An internal thermal sensor controlling temperature preference in *Drosophila*. *Nature* **454**, 217–220 (2008).
- Gallio, M., Ofstad, T. A., Macpherson, L. J., Wang, J. W. & Zuker, C. S. The coding of temperature in the *Drosophila* brain. *Cell* **144**, 614–624 (2011).
- Thorne, N., Chromey, C., Bray, S. & Amrein, H. Taste perception and coding in *Drosophila*. *Curr. Biol.* **14**, 1065–1079 (2004).
- Vosshall, L. B. & Stocker, R. F. Molecular architecture of smell and taste in *Drosophila*. *Annu. Rev. Neurosci.* **30**, 505–533 (2007).
- Silbering, A. F. & Benton, R. Ionotropic and metabotropic mechanisms in chemoreception: 'chance or design'? *EMBO Rep.* **11**, 173–179 (2010).
- Viswanath, V. et al. Opposite thermosensor in fruitfly and mouse. *Nature* **423**, 822–823 (2003).
- Foelix, R. F., Stocker, R. F. & Steinbrecht, R. A. Fine structure of a sensory organ in the arista of *Drosophila melanogaster* and some other dipterans. *Cell Tissue Res.* **258**, 277–287 (1989).
- Thorne, N. & Amrein, H. Atypical expression of *Drosophila* gustatory receptor genes in sensory and central neurons. *J. Comp. Neurol.* **506**, 548–568 (2008).
- Robertson, H. M., Warr, C. G. & Carlson, J. R. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **100** (suppl. 2), 14537–14542 (2003).
- Sayed, O. & Benzer, S. Behavioral genetics of thermosensation and hyposensation in *Drosophila*. *Proc. Natl Acad. Sci. USA* **93**, 6079–6084 (1996).
- Sweeney, S. T., Broadie, K., Keane, J., Niemann, H. & O'Kane, C. J. Targeted expression of tetanus toxin light chain in *Drosophila* specifically eliminates synaptic transmission and causes behavioral defects. *Neuron* **14**, 341–351 (1995).
- Zars, T. Two thermosensors in *Drosophila* have different behavioral functions. *J. Comp. Physiol. A* **187**, 235–242 (2001).
- Kim, S. H. et al. *Drosophila* TRPA1 channel mediates chemical avoidance in gustatory receptor neurons. *Proc. Natl Acad. Sci. USA* **107**, 8440–8445 (2010).
- Kang, K. et al. Modulation of TRPA1 thermal sensitivity enables sensory discrimination in *Drosophila*. *Nature* **481**, 76–80 (2012).
- Pulver, S. R., Pashkovski, S. L., Hornstein, N. J., Garrity, P. A. & Griffith, L. C. Temporal dynamics of neuronal activation by Channelrhodopsin-2 and TRPA1 determine behavioral output in *Drosophila* larvae. *J. Neurophysiol.* **101**, 3075–3088 (2009).
- Vyklický, L. et al. Temperature coefficient of membrane currents induced by noxious heat in sensory neurones in the rat. *J. Physiol. (Lond.)* **517**, 181–192 (1999).
- Bernstein, J. G., Garrity, P. A. & Boyden, E. S. Optogenetics and thermogenetics: technologies for controlling the activity of targeted cells within intact neural circuits. *Curr. Opin. Neurobiol.* **22**, 61–71 (2012).
- Gingl, E., Hinterwirth, A. & Tichy, H. Sensory representation of temperature in mosquito warm and cold cells. *J. Neurophysiol.* **94**, 176–185 (2005).
- Sato, K., Tanaka, K. & Touhara, K. Sugar-regulated cation channel formed by an insect gustatory receptor. *Proc. Natl Acad. Sci. USA* **108**, 11680–11685 (2011).
- Liu, J. et al. *Drosophila* phototransduction requires a G protein-dependent cAMP pathway and a taste receptor homolog. *Nature Neurosci.* **13**, 715–722 (2010).
- Xiang, Y. et al. Light-avoidance-mediating photoreceptors tile the *Drosophila* larval body wall. *Nature* **468**, 921–926 (2010).
- Wang, G. et al. *Anopheles gambiae* TRPA1 is a heat-activated channel expressed in thermosensitive sensilla of female antennae. *Eur. J. Neurosci.* **30**, 967–974 (2009).

25. Cho, H. *et al.* The calcium-activated chloride channel anoctamin 1 acts as a heat sensor in nociceptive neurons. *Nature Neurosci.* **15**, 1015–1021 (2012).
26. Xiao, B., Coste, B., Mathur, J. & Patapoutian, A. Temperature-dependent STIM1 activation induces Ca^{2+} influx and modulates gene expression. *Nature Chem. Biol.* **7**, 351–358 (2011).
27. Kang, K. *et al.* Analysis of *Drosophila* TRPA1 reveals an ancient origin for human chemical nociception. *Nature* **464**, 597–600 (2010).
28. Suchard, M. A. & Redelings, B. D. BAli-Phy: simultaneous Bayesian inference of alignment and phylogeny. *Bioinformatics* **22**, 2047–2048 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the Garrity laboratory, C. B. Chien, L. Huang, R. Huey, E. Marder and G. Turrigiano for comments, H. Amrein, Y. Jan, C. Montell, C. Zuker and Bloomington Stock Center for strains, N. Donelson, M. Klein, A. Samuel, T. Lauer, P. Taneja, S. Nelson, Y. Yu, H. Bell, P. Sengupta, F. Baier and L. Vossahl for assistance. Supported by grants from National Institute of Mental Health (NIMH) (EUREKA R01

MH094721), National Institute of Neurological Disorders and Stroke (NINDS) (P01 NS044232) and National Science Foundation (IOS-1025307) to P.A.G., National Institute of General Medical Sciences (NIGMS) (R01 GM054408) and NIMH (R01 MH067284) to L.C.G., NIGMS (R01 GM094468) to D.L.T. and NINDS National Research Service Award to V.C.P.

Author Contributions L.N., P.B., A.M.L., L.C.G. and P.A.G. designed experiments. L.N. performed molecular genetics, behaviour and chemosensor electrophysiology. P.B. performed neuromuscular electrophysiology. E.C.C., A.M.L. and J.O.F. performed behaviour electrophysiology. V.C.P. performed chemosensor electrophysiology. D.L.T. and P.A.G. performed bioinformatics. L.N., P.B., D.L.T., L.C.G. and P.A.G. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.A.G. (pgarrity@brandeis.edu).

METHODS

Fly strains. *Gr28b*, *TrpA1*, *HC-GAL4* and *CC-GAL4* strains were previously described^{3,4,10,15,16,23,27}. *Df(Gr28b)* is *Df(2L)Exel7031* (Bloomington Stock Center). To control for transposon position effects, all *UAS-Gr28* transgenes were inserted at the same landing site, *atp2*, by site-specific transgenesis¹⁶. *UAS-Gr28b(B)* was created from expressed sequence tag clone IP03356 (DGRC stock no. 1623277). Alternative amino termini of *UAS-Gr28b(A)*, *UAS-Gr28b(C)* and *UAS-Gr28b(D)* were amplified from cDNA with N-terminal primers (*UAS-Gr28b(A)*: 5'-CCG GAATTCATGATCCGCTGCGGATTGGAC-3'; *UAS-Gr28b(C)*: 5'-CCGGAAT TCATGGACATTGAAATGGCCAAGG-3'; and *UAS-Gr28b(D)*: 5'-CCGGAA TTCATGTCATTTTACTTTTGGCGAA-3') and common primer (5'-TCCGCAG GATCCTTGGTTACAATGG-3'). *UAS-Gr28b(E)* was amplified from genomic DNA with primers 5'-CCGGAATTCATGTGGCTCCTTAGGCGATCGG-3' and 5'-TCCGCAGGATCCTTGGTTACAATGG-3'. The first intron of *Gr28b(E)* was deleted by PCR (5'-GCACTTAACGAGGTGTGAAGAACC-3' and 5'-G GTTCTTCAACACCTCGTTAAGTGC-3'). *UAS-Gr28a* transgene was amplified from genomic DNA with primers 5'-CCGGAATTCATGGCCTTTAAGTTGTGG GAG-3' and 5'-TCCCCTCGAGGTATATATAATTTAATCAATCG-3'. The introns were deleted by PCR (first intron: 5'-TATCCTGCAGGATTTTCGTTAAAC ATACTAA-3' and 5'-TTAGTATGTTAAACGAAATCCTGCAGGATA-3'; second intron: 5'-GGCAGCACCAGTAATCGTAAAAATCAGTGTG-3' and 5'-CA CACTGATTTTACGATTACTGGTGTGCC-3'). All clones were sequenced to confirm that no mutations were introduced. TRPA1(A) resembles dTRPA1-D²⁹, but contains 20 additional N-terminal amino acids. TRPA1(B) corresponds to dTRPA1-A²⁹.

Behaviour and electrophysiology. Two-temperature rapid negative thermotaxis assay was performed as described¹², except that tube temperatures were 25.5 ± 0.3 °C and 31.0 ± 0.5 °C (± s.d.), ≥15 flies per trial. Ablations were performed with Ultra Fine Clipper Scissors (Fine Science Tools) on ice-anaesthetized 1–4-day-old *white Canton-S* flies. Recovery was 1 h to 2 days. For rapid (1 min) phototaxis, all flies were collected after thermotaxis assay and re-tested using same apparatus but clear-walled tubes (BD Falcon) in a dark-lined box exposed on one side to ultraviolet light (ULTRA-LUM, no. UVA-16). Thermal preference assay was performed as described^{3,12}, with 20–60 flies (2–5 days old) per trial. For proboscis extension, female flies (1–5 days old) were starved overnight with water, then glued to glass slides and heated¹⁶. Flies were given three 5-s heat presentations at 5-s intervals.

Extracellular recordings of gustatory neurons were performed as described¹⁶. At least three animals and six bristles were examined per genotype.

Neuromuscular junction potentials were recorded by current clamp from muscle 6 with 3 M KCl-filled intracellular electrodes (20–30 MΩ) in 0.4 mM Ca²⁺ HL3.1, using an Axoclamp2B (Molecular Devices) and a Digidata1322A (Molecular Devices), recording at 5 kHz with pClamp8 (Molecular Devices). Muscles had –45 mV or lower resting potentials. Perfusate was heated with SC-20/CL100 cooler/controller (Warner Instruments) and temperature monitored with bath thermistor (Warner Instruments) or IT-23 thermocouple (Physitemp) connected to 80TK Thermocouple (Fluke).

Muscle currents were recorded by two-electrode voltage clamp at –60 mV from muscle 6 as above, but using 0 Ca²⁺ HL3.1 solution with 0.5 mM EGTA, 12 mM MgCl₂, 100 μM quinidine and 1 mM 4-AP, and 3 M KCl-filled voltage-sensing (10–15 MΩ) and 3 M CH₃COOK-filled current-passing (5–10 MΩ) electrodes. For light responsiveness, experiments were as above with 0 Ca²⁺ HL3.1 with 0.5 mM EGTA, 4 mM MgCl₂, 100 μM quinidine and 1 mM 4-AP. 30–50 s dark baseline (<0.1 μW mm^{–2} at 400 nm) was recorded, followed by two ~30-s pulses from halogen source at indicated intensities, followed by heat ramp. Intensity was

measured using PM100 light meter (Thor) with 400 nm wavelength correction. Intensity (in mW mm^{–2}) across wavelengths measured: pulse 1: 1.4 at 400 nm, 0.25 at 488 nm, 0.17 at 577 nm, 0.02 at 700 nm; pulse 2: 4.3 at 400 nm, 1.02 at 488 nm, 0.75 at 577 nm, 0.1 at 700 nm. Intensities are minimum estimates; meter was ~2 mm further from source than preparation.

Motor neuron currents were recorded at –60 mV by whole-cell patch clamp with Multiclamp700A amplifier (Molecular Devices) and patch pipettes (3.5–4 MΩ). External solution was a nominally Ca²⁺-free modified A solution (in mM: 118 NaCl, 2 KCl, 4 MgCl₂, 5 Trehalose, 45.5 sucrose, 5 HEPES), 290 mOsm, pH 7.1–7.2, with 0.15 μM tetrodotoxin to limit spiking. The internal solution (in mM: 2 NaCl, 130 K-gluconate, 0.1 CaCl₂, 2 MgCl₂, 1 EGTA, 0.2 Na-GTP, 10 HEPES) adjusted to 285 mOsm with glucose, and pH 7.1–7.2 with KOH. Dorsal motor neurons below nerve cord sheath were visualized with ×40 water immersion objective and exposed using 0.75% (w/v) protease (type XIV, Sigma) in modified A solution. For ion substitution, after initial heating, perfusion was changed to nominally Ca²⁺-free external Modified A solution of the same osmolarity with NaCl replaced by equimolar NMDG and HCl. After 5 min NMDG solution perfusion, preparation was reheated. Perfusion was then reverted to nominally Ca²⁺-free modified A solution. After 5 min, a third heat ramp was recorded. Trace plotting and analysis performed in Matlab. All neuromuscular physiology used female third instar larvae.

The data presented reflect biological replicates as noted in each sample's *n*. Sample sizes were chosen to reliably reveal robust distinctions among samples. No blinding or randomization was used. Nonparametric analysis (Kruskal–Wallis/Steel–Dwass All Pairs test (JMP10, SAS)) yielded results similar to Tukey's HSD.

Molecular biology. Transgenic fly creation and RT-PCR were performed as described¹⁶. RT-PCR primers straddled splice junctions to minimize genomic DNA amplification. Three independent tissue preparations gave similar results. Primers for RT-PCR reactions: *Gr28a* forward primer: 5'-CAGCACCAGTA ATCGTAAAAATC-3'; *Gr28a* reverse primer: 5'-TATGTTAAACGAAATCC TGCAGG-3'; *Gr28b(A)* forward primer: 5'-AACGTTTGCGAAGTCCTGTGTC-3'; *Gr28b(B)* forward primer: 5'-GCTGTGATTTATACGTCCGC-3'; *Gr28b(C)* forward primer: 5'-CTGTCATCTACCTGACTGCC-3'; *Gr28b(D)* forward primer: 5'-TTCCTCTGCAGCAGCATTCG-3'; *Gr28b(A)*, *Gr28b(B)*, *Gr28b(C)* and *Gr28b(D)* common reverse primer: 5'-TCCTGTATAATCTCCGAGG-3'; *Gr28b(D)* reverse primer (used in Supplementary Fig. 1a): 5'-CTTGACCTCAGCACITTTGG-3'; *Gr28b(E)* forward primer: 5'-GGCCCCGTGATCGTGAAA-3'; *Gr28b(E)* reverse primer: 5'-GCACTTAACGAGGTGTGAAG-3'.

Phylogeny. As gustatory receptor sequence diversity creates the potential for alignment ambiguities, phylogeny was created using Bali-Phy²⁸, which performs simultaneous Bayesian inference of alignment and phylogeny. The analysis was performed using the RS07 insertion/deletion model³⁰, LG substitution matrix³¹, estimating equilibrium amino acid frequencies, with gamma distributed rate variation (four categories). Two independent chains were run until the average standard deviation of split frequencies (ASDSF) and potential scale reduction factor (PSRF) based on width of 80% credible intervals (PSRF-80%CI) criteria fell below 0.01 and 1.01, respectively.

29. Zhong, L. *et al.* Thermosensory and nonthermosensory isoforms of *Drosophila melanogaster* TRPA1 reveal heat-sensor domains of a thermoTRP channel. *Cell Rep.* **1**, 43–55 (2012).
30. Redelings, B. D. & Suchard, M. A. Incorporating indel information into phylogeny estimation for rapidly emerging pathogens. *BMC Evol. Biol.* **7**, 40 (2007).
31. Le, S. Q. & Gascuel, O. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25**, 1307–1320 (2008).

Dietary intervention impact on gut microbial gene richness

Aurélien Cotillard^{1,2*}, Sean P. Kennedy^{3*}, Ling Chun Kong^{1,2,4*}, Edi Prifti^{1,2,3*}, Nicolas Pons^{3*}, Emmanuelle Le Chatelier³, Mathieu Almeida³, Benoit Quinquis³, Florence Levenez^{3,5}, Nathalie Galleron³, Sophie Gougis⁴, Salwa Rizkalla^{1,2,4}, Jean-Michel Batto^{3,5}, Pierre Renault⁵, ANR MicroObes consortium†, Joel Doré^{3,5}, Jean-Daniel Zucker^{1,2,6}, Karine Clément^{1,2,4} & Stanislav Dusko Ehrlich³

Complex gene–environment interactions are considered important in the development of obesity¹. The composition of the gut microbiota can determine the efficacy of energy harvest from food^{2–4} and changes in dietary composition have been associated with changes in the composition of gut microbial populations^{5,6}. The capacity to explore microbiota composition was markedly improved by the development of metagenomic approaches^{7,8}, which have already allowed production of the first human gut microbial gene catalogue⁹ and stratifying individuals by their gut genomic profile into different enterotypes¹⁰, but the analyses were carried out mainly in non-intervention settings. To investigate the temporal relationships between food intake, gut microbiota and metabolic and inflammatory phenotypes, we conducted diet-induced weight-loss and weight-stabilization interventions in a study sample of 38 obese and 11 overweight individuals. Here we report that individuals with reduced microbial gene richness (40%) present more pronounced dys-metabolism and low-grade inflammation, as observed concomitantly in the accompanying paper¹¹. Dietary intervention improves low gene richness and clinical phenotypes, but seems to be less efficient for inflammation variables in individuals with lower gene richness. Low gene richness may therefore have predictive potential for the efficacy of intervention.

To examine relationships between variations in gut microbiota composition and bioclinical parameters after dietary intervention, we used the approach termed quantitative metagenomics¹¹. Forty-nine obese or overweight subjects were recruited and subjected to a 6-week energy-restricted high-protein diet followed by a 6-week weight-maintenance diet (Methods); the compliance was good, as indicated by a principal component analysis (PCA) of 35 nutrients over time (Supplementary Fig. 1). Bioclinical characteristics and detailed qualitative and quantitative features of individuals' food intake were obtained at baseline, 6 and 12 weeks (Supplementary Tables 1 and 2). The 35% decrease in energy intake after the first 6 weeks was associated with a reduction in body-fat mass, adipocyte diameter and improvements in insulin sensitivity and markers of metabolism and inflammation (Supplementary Tables 1 and 3). During the weight-maintenance phase, intake of nutrients tended to return to baseline values, whereas dietary total energy, carbohydrate and lipid intake remained lower than at beginning of the intervention (Supplementary Tables 2 and 3). Serum lipid variables also tended to return to their basal levels as well, while a progressive reduction occurred in systemic inflammation markers.

We first examined the gut microbial composition of the study population at baseline (Methods). A bimodal distribution of bacterial gene number was observed (Fig. 1a), similar to the one found in a cohort of 292 Danish individuals¹¹, albeit less distinct, possibly owing to a lower

cohort size. At a threshold of 480,000 genes, corresponding to that from the accompanying manuscript¹¹, there were 18 (40%) low gene count (LGC) and 27 (60%) high gene count (HGC) individuals, harbouring on average 379,436 and 561,499 genes respectively, a one-third difference. A difference in diversity between lean and obese individuals was reported previously¹², but the difference among the obese was not described.

We then examined the baseline phenotypes of the study population. The LGC group had significantly higher insulin resistance and fasting serum triglyceride levels, as well as a tendency towards higher LDL cholesterol and inflammation than the HGC group (Fig. 2); as observed in the accompanying paper¹¹. Analysing gene richness as a quantitative variable gave similar results (Supplementary Table 4). We conclude that in two European countries, the individuals of the LGC group present phenotypes that expose them to an increased risk of obesity-associated co-morbidities. Antibiotic treatments, which lower the diversity, have

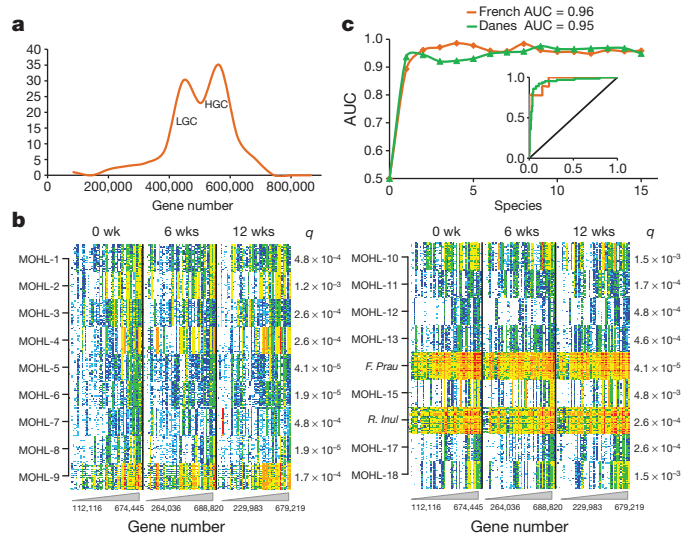


Figure 1 | Gut microbial composition of LGC ($n = 18$) and HGC ($n = 27$) subjects. **a**, Baseline gene count. **b**, Presence and frequency of 25 tracer genes for species differentially abundant in LGC and HGC groups; Mann–Whitney probability (q , false discovery rate (FDR) adjusted) is given. Genes are in rows, frequency is indicated by colour gradient (white, not detected; red, most abundant); individuals, ordered by increasing gene number, are in columns. **c**, Highest AUC values for a combination of a given number of species in a ROC analysis of 45 individuals of our cohort (red) and 292 individuals of the Danish cohort¹¹. Inset, AUC for the combination of six species.

¹Institut National de la Santé et de la Recherche Médicale, U872, Nutrimique, Équipe 7, Centre de Recherches des Cordeliers, Paris 75006, France. ²Université Pierre et Marie-Curie-Paris 6, Nutrimique, 15 rue de l'École de Médecine, Paris 75006, France. ³INRA, Institut National de la Recherche Agronomique, Metagenopolis, Jouy en Josas 78350, France. ⁴Institute of Cardiometabolism and Nutrition, Assistance Publique-Hôpitaux de Paris, CRNH-Ile de France, Pitié-Salpêtrière, Boulevard de l'Hôpital, Paris 75013, France. ⁵INRA, Institut National de la Recherche Agronomique, UMR 1319 Micalis, Jouy en Josas 78350, France. ⁶Institut de Recherche pour le Développement, IRD, UMI 209, UMMISCO, France Nord, Bondy F-93143, France.

*These authors contributed equally to this work.

†A list of authors and affiliations appears at the end of the paper.

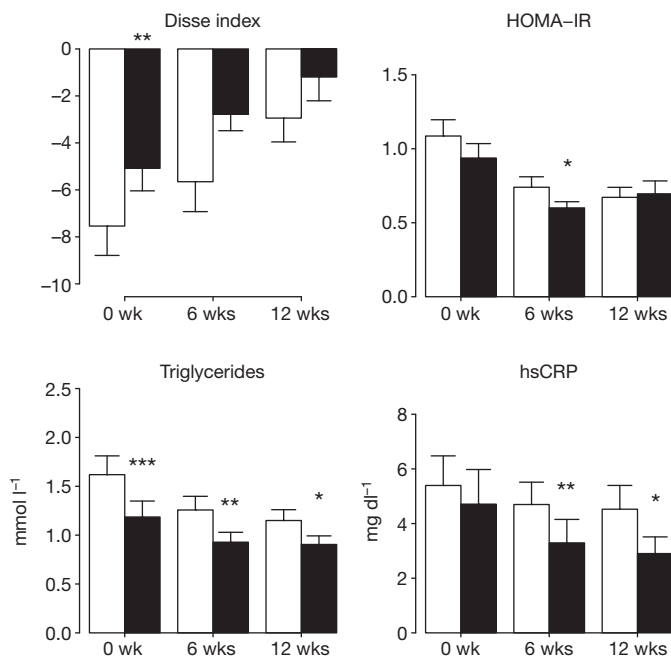


Figure 2 | Differences between LGC and HGC subjects in bioclinical variables. White and black bars refer to LGC ($n = 18$) and HGC ($n = 27$) groups, respectively; error bars denote s.e.m. 0 weeks, baseline; 6 weeks, end of the energy restriction period; and 12 weeks, end of stabilization period. * $P < 0.1$, ** $P < 0.05$, *** $P < 0.01$ by Mann–Whitney tests. ‘Disse index’ is calculated by combining lipid and insulin values (see Supplementary Information). HOMA-IR, homeostatic model assessment of insulin resistance; hsCRP, highly sensitive C-reactive protein.

been reported to improve the hormonal, metabolic and inflammatory status of obese mice; this apparent contradiction may be due to a restoration of a balance of the pro-inflammatory and inflammatory bacterial species in mice. Interestingly, LGC subjects seemed to consume less fruits and vegetables and less fishery products than HGC subjects (Supplementary Tables 4–6), raising the possibility that long-term dietary habits may affect gene richness and the associated phenotypes, as suggested for the elderly¹³.

We next searched for bacterial species differentially abundant in the LGC and the HGC groups. To this aim, we first identified the genes that had significantly different frequencies in the LGC and HGC groups and then clustered the genes supposedly from the same species by a frequency-based covariance analysis (Methods). We identified 6,230 genes that were different according to a Mann–Whitney test ($P < 0.0001$); 4,462 (72%) were grouped into 112 clusters containing at least 2 genes with a Spearman correlation coefficient $\rho > 0.85$. A vast majority of these genes (3,966; 89%) were found in only 18 clusters, which originate from species differentially abundant in the LGC and HGC groups (Supplementary Table 7). The relative abundance of the 18 clusters in each individual was computed as a mean frequency of the 25 tracer genes for each cluster; all were significantly more abundant among the HGC individuals (Fig. 1b and Supplementary Table 8).

To test whether the LGC and HGC individuals could be distinguished using the 18 species represented by the tracer genes, we carried out an exhaustive receiver operating characteristic (ROC) analysis of all clusters combinations, with tenfold cross validation, using 90% of individuals for computation and the remaining 10% for test (Methods). The best area under the curve (AUC) values for combination of different numbers of species are shown in Fig. 1c; they ranged between 0.96 and 0.99 for 2 to 9 species combinations, indicating an almost perfect stratification of LGC and HGC individuals. Interestingly, 14 of the 18 species represented by the tracer genes (78%) were also identified as differentially abundant among the LGC and HGC individuals in a larger Danish cohort¹¹. Not surprisingly, the combinations yielding

the best AUC values for our cohort also efficiently stratified LGC and HGC Danes (Fig. 1c). This indicates that the LGC and HGC individuals from two European countries differ in a similar way, not only by their clinical phenotypes but also by specific features of their gut microbiota.

Very interestingly, gene richness increased significantly in the LGC group after the energy-restricted diet and remained after the stabilization phase higher than at baseline even though a slight downwards trend was apparent, whereas it did not change significantly during intervention in the HGC group (Fig. 3a). We conclude that a dietary intervention can correct a putative loss of richness in the LGC group, albeit partially, as the difference between the LGC and HGC groups remained significant at the end of the intervention.

To investigate the potential effect of the increase in gene richness on patient status we analysed association of the changes of richness and of bioclinical variables. Increase of gene richness was associated with a significant decrease in adiposity measures (hip circumference and total fat mass) and circulating cholesterol as well as a trend towards a decrease in inflammation (highly sensitive C-reactive protein) (Supplementary Table 9). These results suggest that the correction of a putative loss of microbial richness is associated with an improvement of the systemic metabolic status. However, although the inflammation was decreased in all individuals, the difference between LGC and HGC individuals was not attenuated (Fig. 2). Low basal gene richness was also associated with increased adipose tissue inflammatory cells at 6 weeks and increased

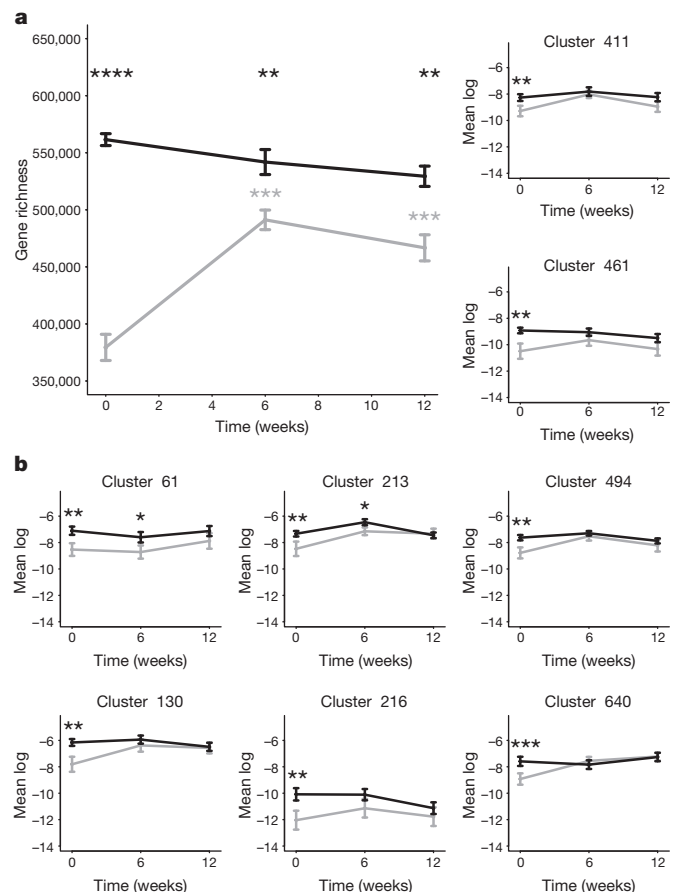


Figure 3 | Gene richness of LGC and HGC groups during the intervention. Data are mean \pm s.e.m. Black line, HGC ($n = 27$); grey line, LGC ($n = 18$). Differences between HGC and LGC groups were tested using Mann–Whitney tests (black asterisks). * $P < 0.1$, ** $P < 0.05$, *** $P < 0.01$, **** $P < 0.001$. a, Overall pattern of variation. For each group, differences between one time point and basal state were tested using Wilcoxon signed-rank tests (grey asterisks). b, Variation of eight clusters that were significantly different at baseline and modulated by the dietary intervention.

systemic inflammation at 12 weeks (Supplementary Table 4). Furthermore, higher gene richness at baseline was associated with a more marked improvement of adipose tissue and systemic inflammation (delta changes at 6 and 12 weeks, respectively; Supplementary Table 10). Gene richness may therefore help to predict the efficacy of dietary intervention on inflammatory variables in overweight or obese individuals.

To further explore the effects of dietary intervention on gut microbial species we used a gene clustering procedure similar to the one described above for the comparison of LGC and HGC individuals. A set of 213,532 genes that varied significantly in frequency between different time points (Wilcoxon signed-rank test, $P < 0.05$) was first identified. To reduce the complexity of the data set, an entropy-filtering criterion was then applied, removing the genes present in only a few individuals (Supplementary Fig. 2). The remaining 58,109 genes were clustered by frequency covariance (Methods and Supplementary Fig. 3). Some 34,920 genes (60%) were grouped in 39 clusters larger than 100 genes (Supplementary Table 11); a large majority, 72%, were very compact, with a clustering coefficient >0.75 (ref. 14) (Supplementary Information, see cluster sheets for a more detailed description). Of the 39 clusters, 17 had $\geq 80\%$ of their genes assigned to the same species and 19 to the same genus (the global distribution was 64% Firmicutes, 33% Bacteroidetes, and 3% Actinobacteria; Supplementary Table 11), confirming a species-specific clustering (Methods).

The abundance of the potential species represented by the 39 clusters was computed as the sum of the respective gene frequencies, and variations over time and correlations with bioclinical variables and food items were examined (Methods and Supplementary Tables 12 and 13). We observed that the abundance of 26 clusters varied significantly with time, indicating that a number of bacterial species can be modulated by nutritional intervention; the remaining 13 were not studied further. Only a few of our gene clusters decreased or showed a tendency to decrease during the calorie restriction phase, but one of those was assigned to *Eubacterium rectale* and another one to *Bifidobacterium* spp., in accordance with previous results⁶ (Supplementary Table 11 and Supplementary Information).

The main trend after 6 weeks of energy-restricted diet was a significant increase of abundance of most gene clusters ($n = 15$), whereas the trend was opposite after 6 weeks of weight-maintenance diet, as the abundance of 14 species decreased. A total of five different patterns was observed, reflecting combinations of variation during the two periods (Supplementary Table 11), but the overall tendency was to return close to a baseline level by the end of the weight-maintenance phase (illustrated in Supplementary Information), suggesting a transient effect of dietary intervention on gut microbiota, as described previously¹⁵. Interestingly, for 8 of the 26 gene clusters that had a significantly lower abundance in the LGC than HGC individuals at baseline (Supplementary Table 14), the energy-restricted diet led to an increase of abundance in the LGC individuals, bringing them close to the level found in the HGC individuals (Fig. 3b); there was no significant abundance difference between the LGC and HGC individuals upon the stabilization phase. We conclude that the dietary intervention, in spite of its overall transient effect, may lead to more persistent changes of some gut microbial species.

Quantitative metagenomics analysis of the gut microbiome in 3 different samples for each of the 49 French (our study) and in 292 Danish subjects¹¹ revealed the existence of a high proportion of individuals (23–40%) with low microbial richness. In both study populations, a detailed clinical analysis indicated that these individuals show adiposity associated dyslipidaemia, higher insulin resistance and low-grade inflammation when compared to their higher-gene-diversity counterparts. This deleterious phenotype is known to be associated with increased risk of pre-diabetes, type 2 diabetes, hepatic and cardiovascular disorders as well as some forms of cancer^{16–18}. In both study populations, abundance of many gut bacterial species in low-richness individuals was altered in a similar way relative to high richness individuals; this alteration can be accurately detected by combinations of only a few bacterial species. This indicates that simple diagnostic tests, based on

our ‘other genome’ could be developed to identify individuals at a higher risk of obesity-associated co-morbidities. In the context of the current global epidemics of obesity and metabolic disorders, such tests could have a broad usefulness.

The concomitant improvement of gut microbial gene richness and bioclinical variables in LGC individuals by a dietary intervention suggests a possibility to advance from risk detection to risk alleviation, under the assumption that the less rich microbiota are also less healthy (see the accompanying paper¹¹). Restoration of gene richness was not achieved fully by our short-term intervention, but seems to be a desirable goal, as decreased gene richness was found to be associated with a less efficient improvement of the inflammatory variables by dietary intervention. Interestingly, increased consumption of fruits and vegetable and thus higher fibre consumption before the intervention seemed to be associated with high bacterial richness. This finding, although exploratory in nature and requiring replication, supports a recently reported link between long-term dietary habits and the structure of gut microbiota¹⁵ and suggests that a permanent change of microbiota may be achieved by appropriate diet. Development of a two-pronged approach, coupling early detection of an impending loss of gut bacterial richness to appropriate nutritional recommendations, which is yet to be established, may help to reach this goal and possibly contribute to diminish the risk of the obesity-linked co-morbidities; stratification by gene richness may have predictive value in respect to the efficacy of a dietary treatment and even guide its choice. However, low-grade inflammation, an important trait related to obesity but also common to many chronic diseases, seemed relatively refractory to dietary intervention in the LGC individuals, suggesting that specific therapeutic actions, aiming at restoring gut microbiota richness and equilibrium in obesity and altered metabolism, may need to be developed as well.

METHODS SUMMARY

Forty-nine obese or overweight subjects were recruited and subjected to a 6-week energy-restricted high-protein diet followed by a 6-week weight-maintenance diet. Bioclinical characteristics, physical activity scores and detailed qualitative and quantitative features of their food intake were obtained at baseline, 6 and 12 weeks (Methods). The clinical trial was registered at <http://www.ClinicalTrials.gov> under study number NCT01314690. The Ethical Committee of Hôtel Dieu Hospital approved the clinical study and all subjects provided written informed consent. Faecal samples were collected at each time point and analysed with the next generation sequencing SOLiD System. After read mapping a frequency table of microbial genes was obtained (Methods).

Two groups of patients with LGC and HGC were defined using the gene-richness distribution. Differences in terms of food, bioclinical variables and gene abundance were identified by standard statistical methods (Methods). Focusing on the dietary intervention and using a multi-criteria selection to narrow down the number of genes to a few thousands, gene clusters of co-varying microbial genes were constructed. These resulting gene clusters were then analysed for changes over time and correlations with bioclinical markers (Methods).

Full Methods and any associated references are available in the online version of the paper.

Received 12 April 2012; accepted 17 July 2013.

1. Mutch, D. M. & Clément, K. Unraveling the genetics of human obesity. *PLoS Genet.* **2**, e188 (2006).
2. Bäckhed, F. *et al.* The gut microbiota as an environmental factor that regulates fat storage. *Proc. Natl Acad. Sci. USA* **101**, 15718–15723 (2004).
3. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).
4. Bäckhed, F., Manchester, J. K., Semenkovich, C. F. & Gordon, J. I. Mechanisms underlying the resistance to diet-induced obesity in germ-free mice. *Proc. Natl Acad. Sci. USA* **104**, 979–984 (2007).
5. Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: human gut microbes associated with obesity. *Nature* **444**, 1022–1023 (2006).
6. Duncan, S. H. *et al.* Reduced dietary intake of carbohydrates by obese subjects results in decreased concentrations of butyrate and butyrate-producing bacteria in feces. *Appl. Environ. Microbiol.* **73**, 1073–1078 (2007).
7. Riesenfeld, C. S., Schloss, P. D. & Handelsman, J. Metagenomics: genomic analysis of microbial communities. *Annu. Rev. Genet.* **38**, 525–552 (2004).
8. National Research Council. *The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet* (The National Academies Press, 2007).

9. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
10. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).
11. Le Chatelier, E. *et al.* Richness of human gut microbiome correlates with metabolic markers. *Nature* <http://dx.doi.org/10.1038/nature12506> (this issue).
12. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457**, 480–484 (2009).
13. Claesson, M. J. *et al.* Gut microbiota composition correlates with diet and health in the elderly. *Nature* **488**, 178–184 (2012).
14. Wasserman, S. & Faust, K. *Social Network Analysis: Methods and Applications*. (Cambridge Univ. Press, 1994).
15. Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105–108 (2011).
16. Ouchi, N., Parker, J. L., Lugus, J. J. & Walsh, K. Adipokines in inflammation and metabolic disease. *Nature Rev. Immunol.* **11**, 85–97 (2011).
17. Shoelson, S. E., Lee, J. & Goldfine, A. B. Inflammation and insulin resistance. *J. Clin. Invest.* **116**, 1793–1801 (2006).
18. Renehan, A. G., Tyson, M., Egger, M., Heller, R. F. & Zwahlen, M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet* **371**, 569–578 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to O. Pedersen (Univ. Copenhagen) for helpful comments on this manuscript and to the MetaHIT consortium for providing the gene profiles of the Danish subjects used to test the ROC models in advance of publication and the DNA samples sequenced on the SOLiD platform for comparison with the Illumina platform used in the accompanying manuscript. We thank C. Baudoin, P. Ancel and V. Pelloux who contributed to the clinical investigation study; S. Fellahi and J.-P. Bastard for analyses of inflammatory markers; D. Bonnefont-Rousselot and R. Bittar for help with the analysis of plasma lipid profile. This work was supported by Agence Nationale de la Recherche (ANR MICRO-Obes, ANR, Nutra2sens, ANR-10-IAHU-05), the Metagenopolis grant ANR-11-DPBS-0001, KOT-Cepodi (Florence Massiera), Danone Research (Damien Paineau) and the associations Fondacoœur, and Louis-Bonduelle. Additional funding came from the European Commission FP7 grant HEALTH-F4-2007-201052 and METACARDIS.

Author Contributions S.D.E., J.D. and K.C. designed the study; S.D.E., J.D., K.C. and P.R. managed the study; K.C. and S.R. designed the clinical research; S.R. and L.C.K. conducted the clinical research and clinical data management; A.C., S.R. and L.C.K. conducted clinical and dietary data analysis; S.G. gave dietary counselling to the patients and carried out analysis of dietary data; F.L. prepared the DNA for sequencing; S.K. managed DNA sequencing, which B.Q. and N.G. carried out; N.P. and J.-M.B. established the sequence analysis pipeline; A.C., J.-D.Z., E.P., N.P., E.L.C., M.A., J.-M.B., S.K. and S.D.E. carried out microbial data analysis; A.C., K.C., L.C.K. and S.D.E. wrote the manuscript.

Author Information The raw solid read data for all samples has been deposited in the European Bioinformatics Institute (EBI) European Nucleotide Archive (ENA) under the accession number ERP003699. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.D.E. (dusko.ehrlich@jouy.inra.fr) or K.C. (karine.clement@psl.aphp.fr).

ANR MicroObes consortium members

Hervé Blottière^{1,2}, Marion Leclerc¹, Catherine Juste¹, Tomas de Wouters¹, Patricia Lepage¹, Charlene Fouqueray¹, Arnaud Basdevant³, Cornéliu Henegar³, Cindy Godard³, Marine Fondacci³, Alili Rohia³, Froogh Hajdouch³, Jean Weissenbach⁴, Eric Pelletier⁴, Denis Le Paslier⁴, Jean-Pierre Gauchi⁵, Jean-François Gibrat⁶, Valentin Loux⁶, Wilfrid Carré⁶, Emmanuelle Maguin¹, Maarten van de Guchte¹, Alexandre Jarret¹, Fouad Boumezeur¹ & Séverine Layec¹

¹INRA, Institut National de la Recherche Agronomique, UMR 1319 Micalis, Jouy en Josas 78350, France. ²INRA, Institut National de la Recherche Agronomique, Metagenopolis, Jouy en Josas 78350, France. ³Institute of Cardiometabolism and Nutrition, Assistance Publique-Hôpitaux de Paris, CRNH-Ile de France, Pitié-Salpêtrière, Paris 75013, France. ⁴Commissariat à l'Energie Atomique, Genoscope, Evry 91000, France. ⁵Institut National de la Recherche Agronomique, Mathématiques et Informatique Appliquées, Jouy en Josas 78350, France. ⁶Institut National de la Recherche Agronomique, Mathématique, Informatique et Génome, Jouy en Josas 78350, France.

METHODS

Clinical investigation. Obese ($n = 38$) and overweight ($n = 11$) subjects, 8 men and 41 women, were recruited for a 12-week controlled dietary intervention at the Center of Research in Human Nutrition, Pitié-Salpêtrière Hospital, Paris, France. The subjects included in the study had no chronic pathologies except excess body weight. Their body weight was stable within 3 months before the study. None of the participants was undergoing chronic treatment or had been involved in weight-loss programs in the preceding 12 months. No antibiotics or drugs were taken within 2 months before or during the course of the study. The Ethical Committee of Hôtel Dieu Hospital approved the clinical study and subjects provided written informed consent. In the first 6-week phase, subjects consumed an energy-restricted high-protein diet (1,200 kilocalories (kcal) per day for women and 1,500 kcal for men: 35% proteins, 25% lipids, 44% carbohydrates) with low glycaemic index carbohydrates and enrichment with soluble fibres¹⁹. This phase was followed by a second 6-week body weight stabilization period with 20% increase in total energy intake, above their resting energy metabolic rate. At 0, 6 and 12 weeks, blood and faecal samples were collected and anthropometric measurements were performed. Subjects filled a 7-day dietary record and were interviewed by a registered dietician. On the visit day, the dietician checked the information and clarified any ambiguities regarding detail of food consumed. All records were analysed by the registered dietician using the computer software program PROFILE DOSSIER V3 (Audit Conseil en Informatique Médicale), which has a dietary database initially made up of 400 food items representative of the French diet as described previously²⁰. A nutrient analysis was generated for each subject. Body composition was determined by dual-energy X-ray absorptiometry (DEXA). Blood samples were obtained after 12 h of fasting to measure total cholesterol, high-density lipoprotein (HDL) cholesterol, triglycerides, insulin, glucose, and inflammatory markers (hsCRP and interleukin 6 (IL-6)) as described previously²¹. Insulin resistance was estimated using HOMA-IR and Disse index scores^{22,23}. Subcutaneous abdominal adipose tissue samples were obtained at all time points by needle biopsy from the periumbilical area under local anaesthesia (1% xylocaine) to measure the adipocytes diameter²⁴ and for immunohistochemical studies (HAM56+ -stained macrophages in adipose tissue). Whole faecal samples were self-collected in sterile boxes and stored at -20°C within 4 h, sampled (200-mg aliquots) and then stored at -80°C until analysis. Paired Wilcoxon tests were performed to analyse changes in these variables between various time points ($P < 0.05$). P values were adjusted for multiple testing using the Benjamini–Hochberg procedure²⁵.

Metagenomic sequencing. Intestinal bacterial gene content of 49 obese and overweight individuals at 3 time-points (baseline, week 6 and week 12) was determined by high-throughput ABI SOLiD sequencing technology of total faecal DNA. An average of $76.5 \text{ million} \pm 36.5 \text{ million}$ (mean \pm s.d.) 35-base-long single reads were determined for each sample (a total of 393 Gb of sequence) (Supplementary Table 15). By using corona_lite (v4.0r2.0), an average of $24.8 \text{ million} \pm 14.3 \text{ million}$ reads per individual were mapped on the reference catalogue of 3.3 million genes⁹ with a maximum of 3 mismatches. Reads mapping at multiple positions were discarded and an average of $14.2 \text{ million} \pm 8.1 \text{ million}$ uniquely mapped reads per individuals were retained for estimating the abundance of each reference gene by using METEOR²⁶ software. Abundance of each gene in an individual was normalized with METEOR by dividing the number of reads that uniquely mapped to a gene by its nucleotide length. After that, normalized gene abundances were transformed in frequencies by dividing them with the total number of uniquely mapped reads for a given sample. The resulting set of gene frequencies, termed as a microbial gene profile of an individual, was used for further analyses.

Comparison between SOLiD and Illumina sequencing technologies. Two primary short-read technologies currently exist for quantitative metagenomic analysis; SOLiD and Illumina. To validate data set correspondences and comparisons between results in this study and the accompanying paper¹¹, 24 samples from the Danish Inter99 cohort, previously sequenced on an Illumina GA platform, were also sequenced and analysed by SOLiD technology. Representative samples for cross-comparison included 14 females and 10 males, 15 obese and 9 lean, and 15 HGC and 9 LGC individuals. Hierarchical clustering demonstrated all samples self-clustered as technology-independent pairs, with the average Pearson correlation coefficient of 0.87 (computed upon log transformation) between the two technologies and increasing concordance associated with increased signal (Supplementary Fig. 4).

Gene-richness analysis. Gene richness was compared between subjects using the same number of mapped reads. Data were downsized to adjust for technical variability linked to different sequencing depths. This downsizing was performed at different levels by randomly selecting 4.5 or 7 million mapped reads for each sample and then computing the mean number of genes over 30 drawings (Supplementary Table 15). The 4.5-million-read downsizing allows keeping more than 90% of the individuals at each time point (required for the quantitative analysis of gene richness), but shrinks the data distribution (Supplementary Fig. 5). The 7-million-read downsizing was used for the analysis of the gene count distribution

among the individuals and the enterotypes. The distribution of gene number obtained with the two downsizings is quite similar as shown by Spearman correlation ($\rho > 0.99$) (Supplementary Fig. 5).

Differentially abundant gene clusters between LGC and HGC. Two groups of patients with LGC and HGC were defined using the 480,000-gene threshold, consistent with the accompanying manuscript¹¹ (Fig. 1a, and main text). Genes significantly different in groups of individuals were identified by Mann–Whitney tests using a P -value threshold of < 0.0001 . They were clustered by an abundance-based binning strategy, using the covariance of their gene frequency profiles among the individuals of the cohort, as described in the accompanying paper¹¹. Abundance of a given cluster in each individual was estimated as a mean abundance of 25 arbitrarily selected ‘tracer’ genes for each cluster; these values were close to those obtained by using all the genes of a cluster.

ROC analysis. The analyses were carried out to distinguish between HGC and LGC individuals by a combination of gene clusters. For each combination, only a single decision model was considered, computed as the sum of mean abundance of clusters with greater abundance in HGC than in LGC minus the sum of those with greater abundance in LGC than in HGC. As opposed to the infinite number of regression models, such models are finite and can be exhaustively explored. To select the best models, we used the cross-validated area under the ROC curve cross-validated AUC criterion²⁷ well adapted to classification models for binary outcome data.

Correlations between microbial gene clusters and clinical variables. Mann–Whitney tests were used to compare bioclinical variables, food items and gene clusters between LGC and HGC groups at each time point. Associations between quantitative basal gene richness and bioclinical or food variables, or differences (deltas) in bioclinical or food variables were investigated using linear models. For the associations between deltas of bioclinical parameters and deltas of gene richness, all pairs of deltas were computed (6 weeks–0 weeks, 12 weeks–6 weeks, 12 weeks–0 weeks). Linear mixed models were then fitted using all data. A P -value threshold of 0.05 was applied for statistical significance. Owing to the highly correlated bioclinical and food variables, adjustment for multiple testing is not really adequate, but the false discovery rates (Benjamini–Hochberg²⁵) are given for information purposes in Supplementary Table 9.

Taxonomical annotation. The genes from clusters were mapped by BLASTN (BLAST 2.2.24, default parameters) against a collection of 6,006 genomes (the available reference genomes from NCBI and the set of draft gastrointestinal genomes from the DACC and MetaHIT as of the 03.08.2012). Following taxonomical assignment parameters described by Arumugam¹⁰, each gene was assigned with the taxonomy of the best-hit covering $\geq 80\%$ of the gene length and according to the identity threshold for the taxonomic rank ($\geq 65\%$ for phylum, $\geq 85\%$ for genus and $\geq 90\%$ for species). To assess the taxonomy of clusters below these thresholds we used BLASTP against the non-redundant sequences databases available at NCBI. Based on the criterion of the homogeneity of the best hit taxonomic assignment (at least 80% of tracer genes from a cluster having the same taxonomic best hit assignment), 100% and 25% of the clusters could be assigned at a phylum and genus level, respectively (Supplementary Table 7).

Gene clusters affected by the dietary intervention. The analysis was carried out with genes with a potentially dietary linked signal. The first filtering step consisted in selecting the genes whose frequency was modulated significantly by the nutritional intervention during the dietary restriction or the stabilization period with a Wilcoxon signed-rank test ($P < 0.05$). A subset of these genes, with high Shannon entropy²⁸, was selected in a second filtering step. The entropy distribution of the filtered genes presented a bimodal distribution and the genes corresponding to the highest mode were selected using a threshold estimation on an approximation of its density function²⁹ (Supplementary Fig. 2). The genes with high entropy were mostly shared among individuals of the cohort. Genes with significantly similar frequency profiles (P divided by number of tests < 0.05) and high Spearman correlation coefficient ($\rho > 0.85$), were clustered in a way similar to the LGC–HGC clusters using single-linkage clustering (Supplementary Fig. 3). The 39 clusters with a size superior to 100 genes were kept for further analyses. The group abundance of each cluster was computed as the sum of the frequencies of its genes, and the data were log-transformed for parametric statistics.

Gene-cluster analysis. Gene clusters were analysed for changes over time and correlations with bioclinical markers using linear mixed models were adjusted for age and sex (Supplementary Tables 12 and 13). The highly correlated data induced P values distributions not adapted to standard procedures for multiple testing adjustments; nevertheless, we provide the false discovery rates using the Benjamini–Hochberg method in Supplementary Tables 12 and 14. All statistical analyses were performed using the R environment³⁰.

19. Rizkalla, S. W. *et al.* Differential effects of macronutrient content in 2 energy-restricted diets on cardiovascular risk factors and adipose tissue cell size in

- moderately obese individuals: a randomized controlled trial. *Am. J. Clin. Nutr.* **95**, 49–63 (2012).
20. Bouché, C. *et al.* Five-week, low-glycemic index diet decreases total fat mass and improves plasma lipid profile in moderately overweight nondiabetic men. *Diabetes Care* **25**, 822–828 (2002).
 21. Tordjman, J. *et al.* Structural and inflammatory heterogeneity in subcutaneous adipose tissue: Relation with liver histopathology in morbid obesity. *J. Hepatol.* **56**, 1152–1158 (2012).
 22. Disse, E. *et al.* A lipid-parameter-based index for estimating insulin sensitivity and identifying insulin resistance in a healthy population. *Diabetes Metab.* **34**, 457–463 (2008).
 23. Antuna-Puente, B. *et al.* Evaluation of insulin sensitivity with a new lipid-based index in non-diabetic postmenopausal overweight and obese women before and after a weight loss intervention. *Eur. J. Endocrinol.* **161**, 51–56 (2009).
 24. Prat-Larquemín, L. *et al.* Adipose angiotensinogen secretion, blood pressure, and AGT M235T polymorphism in obese patients. *Obes. Res.* **12**, 556–561 (2004).
 25. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B* **57**, 289–300 (1995).
 26. Pons, N. *et al.* METEOR, a platform for quantitative metagenomic profiling of complex ecosystems. Journées Ouvertes en Biologie, Informatique et Mathématiques <http://www.jobim2010.fr/sites/default/files/presentations/27Pons.pdf> (2010).
 27. Jiang, D., Huang, J. & Zhang, Y. The cross-validated AUC for MCP-logistic regression with high-dimensional data. *Stat. Methods Med. Res.* <http://dx.doi.org/10.1177/0962280211428385> (28 November 2011).
 28. Shannon, C. E. A mathematical theory of communication. *Bell Sys. Tech. J.* **27**, 379–423 (1995), 623–656 (1948).
 29. Silverman, B. W. *Density Estimation for Statistics and Data Analysis* (Chapman and Hall, 1986).
 30. R Development Core Team. *R: A Language and Environment for Statistical Computing* <http://www.R-project.org> (R Foundation for Statistical Computing, 2011).

Coordination of heart and lung co-development by a multipotent cardiopulmonary progenitor

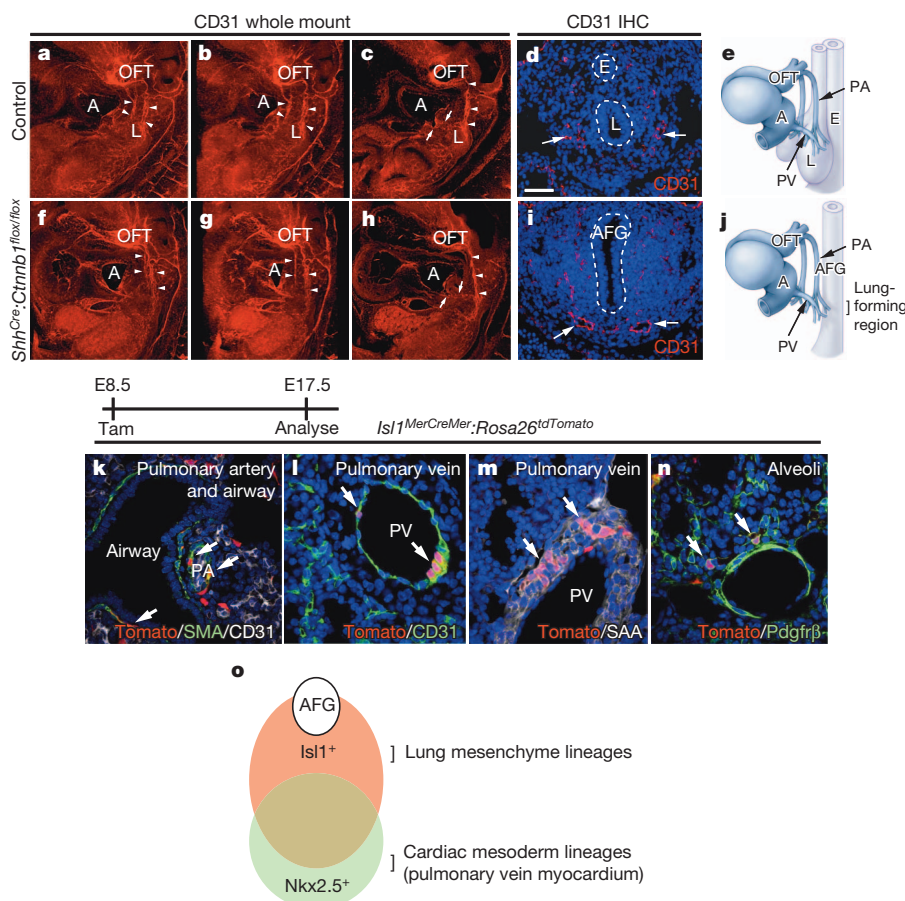
Tien Peng^{1,2}, Ying Tian^{1,3}, Cornelis J. Boogerd⁴, Min Min Lu³, Rachel S. Kadzik⁵, Kathleen M. Stewart^{1,3}, Sylvia M. Evans⁴ & Edward E. Morrisey^{1,3,5,6}

Co-development of the cardiovascular and pulmonary systems is a recent evolutionary adaption to terrestrial life that couples cardiac output with the gas exchange function of the lung¹. Here we show that the murine pulmonary vasculature develops even in the absence of lung development. We have identified a population of multipotent cardiopulmonary mesoderm progenitors (CPPs) within the posterior pole of the heart that are marked by the expression of *Wnt2*, *Gli1* and *Isl1*. We show that CPPs arise from cardiac progenitors before lung development. Lineage tracing and clonal analysis demonstrates that CPPs generate the mesoderm lineages within the cardiac inflow tract and lung including cardiomyocytes, pulmonary vascular and airway smooth muscle, proximal vascular endothelium, and pericyte-like cells. CPPs are regulated by hedgehog expression from the foregut endoderm, which is required for connection of the pulmonary vasculature to the heart. Together, these studies identify a novel population

of multipotent cardiopulmonary progenitors that coordinates heart and lung co-development that is required for adaptation to terrestrial existence.

The coordinated development of the cardiovascular and pulmonary organ systems is illustrated in embryonic development, when the lung endoderm protrudes into the cardiac mesoderm as the two organs develop in parallel to form the cardiopulmonary circulation. However, little is known about the origins and pathways involved in co-development of the cardiopulmonary system. We assessed pulmonary vascular development in a model of lung agenesis through conditional deletion of β -catenin (*Ctnnb1*) within the anterior foregut (AFG) endoderm^{2,3}. The lung fails to develop from the foregut in *Shh^{Cre}:Ctnnb1^{fllox/fllox}* mutant mice, whereas the development of the heart and other foregut-derived organs remains intact (Supplementary Fig. 1)^{2,3}. Using CD31 whole-mount immunostaining and confocal microscopy to visualize early

Figure 1 | The pulmonary vasculature develops in the absence of lung specification. a–o. The paired pulmonary arteries are observed tracking in a posterior direction in both the control (a–c, arrowheads) and *Shh^{Cre}:Ctnnb1^{fllox/fllox}* mutants (f–h, arrowheads) at E10.5 using CD31 whole-mount immunostaining and section immunostaining (d and i, arrows). The primitive pulmonary veins (PVs) are observed emerging from the atria and extending towards the region of the foregut where the lung would normally form in both control and *Shh^{Cre}:Ctnnb1^{fllox/fllox}* mutants (c and h, arrows). These developmental hallmarks are illustrated in models (e and j). *Isl1⁺* progenitors tagged at E8.5 give rise to the pulmonary artery (PA) and airway smooth muscle (k), endothelial (l), myocardial (m), and *Pdgfr β ⁺* pericyte-like (n) lineages of the pulmonary vasculature and other mesodermal derivatives. This overlapping but distinct origin of lung mesoderm derivatives is diagrammed in (o). AT, atrium; AFG, anterior foregut; E, oesophagus; L, lung; OFT, outflow tract; PA, pulmonary artery; PV, pulmonary vein; Tam, tamoxifen. Control, *Shh^{Cre}:Ctnnb1^{fllox/+}* (a–h). Scale bar, 100 μ m.



¹Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ²Division of Pulmonary and Critical Care, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ³Cardiovascular Institute, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ⁴Department of Medicine, Skaggs School of Pharmacy, University of California, San Diego, San Diego, California 92093, USA. ⁵Department of Cell and Developmental Biology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ⁶Institute for Regenerative Medicine, Philadelphia, Pennsylvania 19104, USA.

cardiopulmonary vascular development we show that in wild-type embryos, the paired pulmonary arteries descending from the outflow tract (OFT) of the heart and the pulmonary veins extending from the atria connect with the lung bud to form a vascular plexus at embryonic day 10.5 (E10.5) (Fig. 1a–e and Supplementary Video 1). In *Shh^{Cre};Ctnnb1^{lox/lox}* lung agenesis mutants, the pulmonary arteries and pulmonary veins continue to develop and intersect at a region approximately where the lung bud would normally form (Fig. 1f–i and Supplementary Video 2). Although this vascular plexus persists throughout embryonic development in the absence of lung, it fails to branch or develop further (Supplementary Fig. 2a–d). Interestingly, *Shh* is expressed in the AFG even in the absence of lung development (Supplementary Fig. 2e–h).

The presence of both the pulmonary arteries and the pulmonary veins in the absence of lung development suggests that cardiac progenitors contribute to the generation of these structures. The cardiac progenitors closest to the anterior foregut where the lung arises are defined by expression of *Isl1* and are referred to as the second heart field (SHF) (Supplementary Fig. 3)⁴. *Nkx2.5* expression further subdivides the *Isl1*⁺ domain into a ventral and medial domain that expresses both *Isl1* and *Nkx2.5*, and a lateral and dorsal domain that expresses only *Isl1* (Supplementary Figs 3 and 7). We performed cell-lineage tracing using the *Isl1^{Cre}*, *Isl1^{MerCreMer}* and *Nkx2.5^{Cre}* mouse strains to define the contribution of these lineages to pulmonary mesoderm derivatives (Fig. 1 and Supplementary Fig. 3)^{4,5}. Lineage tracing using the *Isl1^{MerCreMer};Rosa26^{tdTomato}* mice demonstrates that *Isl1*⁺ cells tagged at E8.5 before lung development give rise to all layers of the pulmonary vasculature as well as the myocardium of the cardiac inflow tract (Fig. 1k–n and data not shown). This is supported by lineage tracing with the constitutive *Isl1^{Cre}* demonstrating that *Isl1*⁺ cells generate ventral lung mesenchyme adjoining the inflow tract of the heart, whereas lineage tracing with *Nkx2.5^{Cre}* shows that *Nkx2.5*⁺ progenitors only generate the myocardium of the proximal pulmonary vein (Fig. 1o and Supplementary Fig. 3).

We generated a *Wnt2^{CreERT2}* mouse to delineate the temporal and spatial relationship between the ventral mesoderm flanking the anterior foregut and overlapping the posterior pole of the heart with the development of pulmonary mesoderm lineages (Supplementary Fig. 4). *Wnt2* is expressed in a unique pattern that is confined to the posterior pole of the developing heart adjacent to the AFG before lung development at E8.5 to E9.5 (Fig. 2a, b)^{6,7}. Lineage labelling of *Wnt2*⁺ cells at E8.5 reveals that *Wnt2*⁺ progenitors generate cells within the cardiac inflow tract but not the outflow tract (Supplementary Fig. 5), and these cells move progressively into the developing lung bud (Fig. 2c, d and Supplementary Fig. 6). By E17.5, *Wnt2*⁺ cells labelled at E8.5 generate multiple mesoderm lineages within the developing heart including cardiomyocytes and endocardium (Supplementary Fig. 6). Importantly, *Wnt2*⁺ cardiac progenitors can generate all layers of the pulmonary vasculature, airway smooth muscle, and *Pdgfrβ*⁺/NG2⁺ lung pericyte-like cells similar to the *Isl1*⁺ cardiac progenitor population (Fig. 2e–h and Supplementary Fig. 6). Thus, before lung specification, *Wnt2* marks a multipotent progenitor population that generates the majority of the repertoire of mesodermal lineages in the developing lung and cardiac inflow tract.

Gli1 is expressed in a temporal and spatial pattern similar to that of *Wnt2* and *Isl1* in the ventral mesoderm surrounding the anterior foregut that is activated by *Shh*⁸. We performed lineage tracing with the *Gli1^{CreERT2}* mouse and our data show that *Gli1*⁺ cells marked at E8.5 can contribute to the inflow tract mesoderm of the heart as well as cells surrounding the early lung bud similar to the *Wnt2*⁺ progenitors (Fig. 2i, j and Supplementary Fig. 7). Examination of the fate of E8.5 *Gli1*⁺ progenitors at E17.5 shows that they can generate all of the mesoderm derivatives within the cardiac inflow tract and developing lung including vascular and airway smooth muscle, proximal vessel endothelium, and pulmonary vein myocardium in a manner indistinguishable from *Wnt2*⁺ and *Isl1*⁺ progenitors (Fig. 2k–n and Supplementary

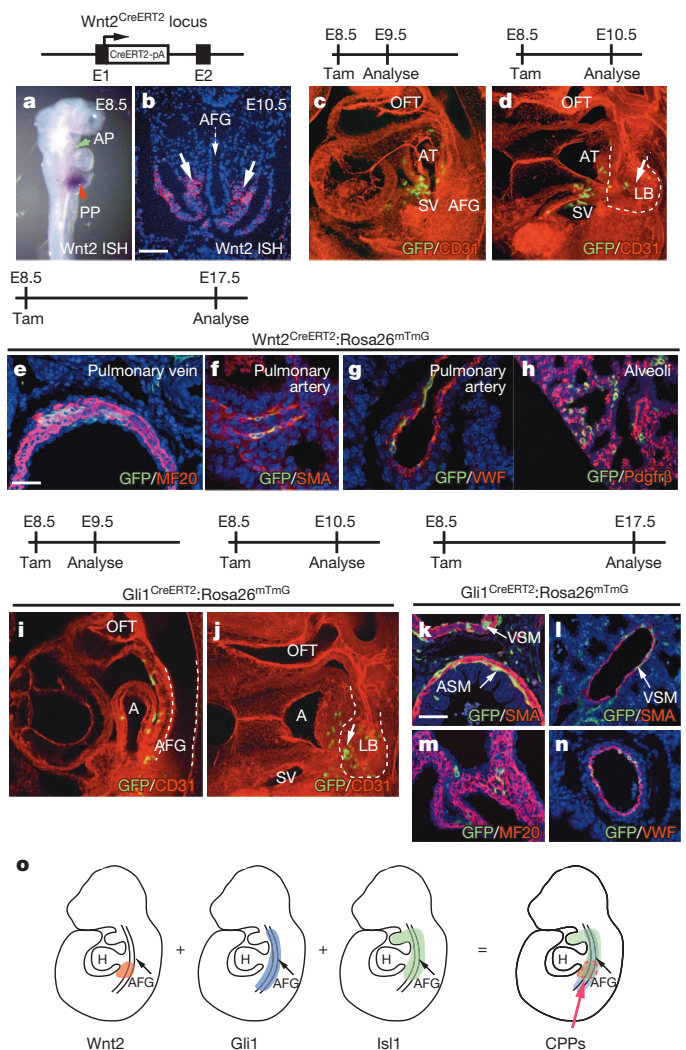


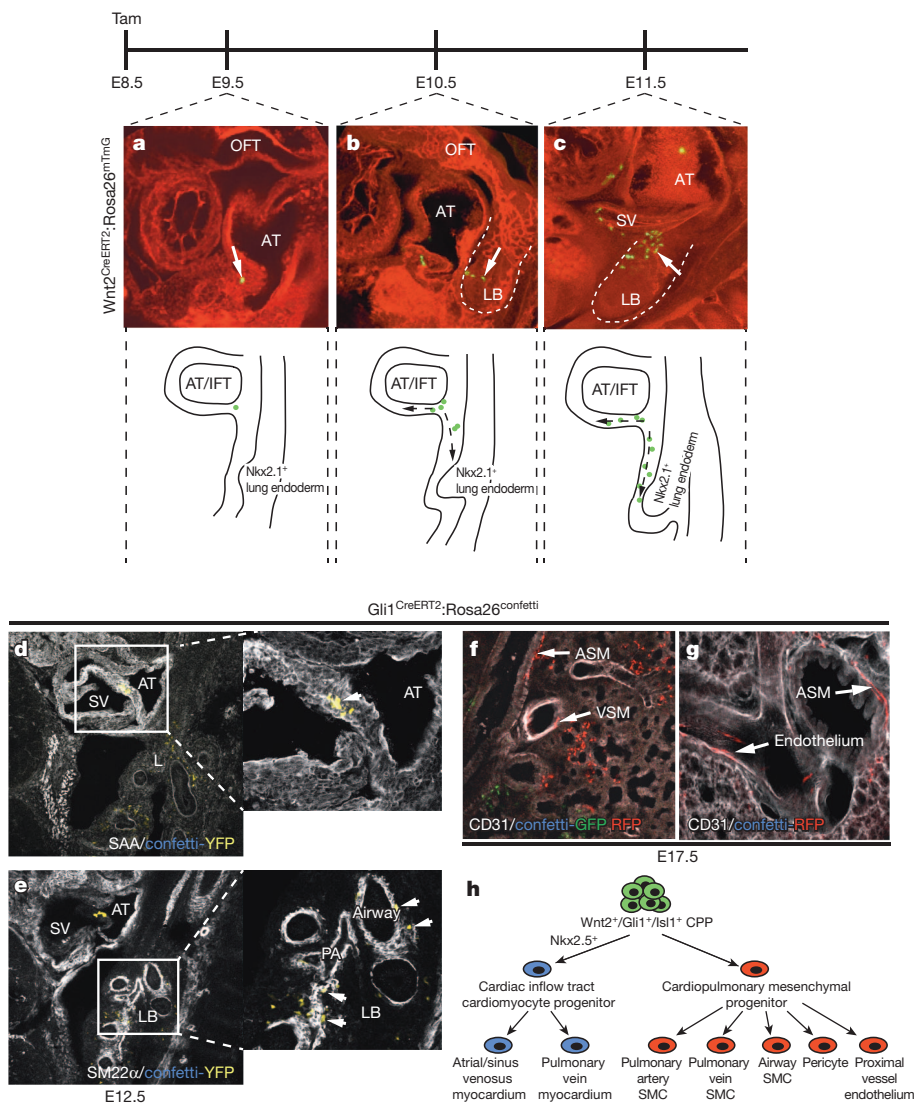
Figure 2 | *Wnt2*⁺ and *Gli1*⁺ cells define a cardiopulmonary progenitor and generate mesoderm derivatives of the lung and cardiac inflow tract.

a–o, *Wnt2* is expressed in the ventral mesoderm surrounding the anterior foregut and overlapping the posterior pole of the heart (**a** and **b**, arrows). *Wnt2*⁺ cells tagged at E8.5 can generate cells within the cardiac inflow tract as well as in the lung mesoderm (**c** and **d**). *Wnt2*⁺ cells tagged at E8.5 generate myocardium of the pulmonary vein, smooth muscle of the pulmonary artery, endothelium of the proximal pulmonary artery, and *Pdgfrβ*⁺ pericyte-like cells in the lung by E17.5 (**e–h**). *Gli1*⁺ cells generate derivatives within the inflow tract and early lung mesoderm similar to *Wnt2*⁺ cells (**i** and **j**) and generate mesoderm lineages within the lung including airway and vascular smooth muscle (**k** and **l**), and endothelium of the proximal pulmonary vessels (**n**) as well as the myocardium of the atria (**m**). CPPs are located in a region of overlapping *Wnt2*, *Gli1* and *Isl1* expression between the developing heart and the anterior foregut (**o**). AP, anterior pole of the heart; ASM, airway smooth muscle; GFP, green fluorescent protein; H, heart; LB, lung bud; PP, posterior pole of the heart; SAA, sarcomeric α -actinin; SMA, smooth muscle actin; SV, sinus venosus; VSM, vascular smooth muscle; VWF, von Willibrand factor. Scale bars, 100 μ m (**b**), 50 μ m (**e–h**), 50 μ m (**k–n**).

Fig. 6). Furthermore, *Wnt2*⁺ progenitors co-express *Isl1* and *Gli1* (Supplementary Fig. 7). These data identify a population of multipotent cardiopulmonary mesoderm progenitors (CPPs) defined by expression of *Wnt2*⁺, *Gli1*⁺ and *Isl1*⁺ that generates the majority of mesoderm lineages in the lung and cardiac inflow tract (Fig. 2o).

Wnt2⁺, *Gli1*⁺ and *Isl1*⁺ progenitors did not contribute in a notable manner to the distal alveolar capillary endothelium of the lung, but did contribute to the VWF⁺ proximal endothelium of the pulmonary vessels (Supplementary Fig. 8). Using the VE-cadherin^{CreERT2}; *Rosa26^{mTmG}*

Figure 3 | Clonal analysis reveals that CPPs generate related lineages within the cardiopulmonary system. **a–h**, Single clones of $Wnt2^{+}$ CPPs contribute to both the developing cardiac inflow tract as well as the mesoderm of the lung (**a–c**). Clonal analysis with the $Gli1^{CreERT2};Rosa26^{confetti}$ line shows that a single CPP clone (yellow fluorescent protein; YFP^{+}) can generate myocardium (SAA^{+}) within the cardiac inflow tract (**d**, arrow) as well as lung mesodermal lineages such as smooth muscle ($SM22\alpha^{+}$) around the airway and PA (**e**, arrows). Clonal analysis of $Gli1^{CreERT2};Rosa26^{confetti}$ lungs show that vascular and airway smooth muscle and endothelium of the proximal vessels share a common $Gli1^{+}$ progenitor at E8.5 (**f** and **g**). A cell-lineage tree showing the relationship of CPPs to differentiated lineages within the cardiopulmonary system (**h**). IFT, inflow tract; RFP, red fluorescent protein.



mouse line, we show that VE-cadherin⁺ endothelial cells at E8.5 give rise to both the proximal and distal alveolar capillary endothelium of the lung⁹ (Supplementary Fig. 8). This suggests that the distal alveolar capillary endothelium arises before lung development from a VE-cadherin⁺ population distinct from CPPs. Interestingly, $Wnt2^{+}$ cells labelled after E12.5 exhibit a statistically significant reduction in their ability to contribute to vascular or airway smooth muscle, the cardiac inflow tract, and proximal vascular endothelium but retain their ability to contribute to the $Pdgfr\beta^{+}$ lung pericyte-like cells in the alveoli (Supplementary Figs 9 and 10).

To define the clonal relationship of cell lineages generated by CPPs, we performed clonal analysis using limiting amounts of tamoxifen at E8.5 in $Wnt2^{CreERT2};Rosa26^{mTmG}$ mice that would reproducibly induce single clones of one to three cells at E9.5 within the posterior pole of the heart (Fig. 3a–c, Supplementary Fig. 11 and Supplementary Tables 1 and 2)¹⁰. All of the clones with cell clusters in the lung had associated clusters in the sinus venosus and the posterior wall of the atria at E10.5 to E11.5, suggesting that CPPs clonally generate cell lineages within the cardiac and pulmonary mesoderm (Fig. 3b, c and Supplementary Table 2). To confirm these results, we performed clonal analysis using $Gli1^{CreERT2};Rosa26^{confetti}$ mice¹¹. Tamoxifen induction at E8.5 led to well-isolated and same-coloured clones located in both the primitive lung bud, the sinus venosus, and the posterior atrial wall (Fig. 3d, e, Supplementary Fig. 11 and Supplementary Table 3). These analyses

also show that vascular smooth muscle, airway smooth muscle, proximal endothelium, and $Pdgfr\beta^{+}$ pericyte-like cells share a common clonal origin in the lung (Fig. 3f–h and Supplementary Table 4).

The expression of *Shh* in the anterior foregut endoderm adjacent to the cardiac inflow tract during pulmonary vascular initiation (Fig. 4a)², and the co-expression of the hedgehog activated effector *Gli1* with *Wnt2* and *Isl1* in CPPs (Supplementary Fig. 7), suggest that hedgehog signalling has an important role in regulating CPP development. *Shh*^{−/−} mutants have a disorganized vascular endothelial plexus along the anterior foregut that fails to connect to the inflow and outflow tract of the heart (Fig. 4b, c). Conditional inactivation of smoothened (*Smo*) within different vascular lineages and in CPPs shows that loss of hedgehog signalling in endothelium and smooth muscle does not affect cardiopulmonary vascular patterning and differentiation (Supplementary Fig. 12)^{12,13}. In contrast, *Smo* inactivation in *Isl1*⁺ CPPs reproduced the *Shh*^{−/−} phenotype (Fig. 4d and Supplementary Fig. 13). *Isl1*^{Cre}.*Smo*^{flx/flx} embryos exhibit a persistence of an aortopulmonary collateral circulation in addition to the persistent truncus arteriosus previously reported (Supplementary Fig. 13)¹⁴. *Isl1*^{Cre}.*Smo*^{flx/flx} mutants also demonstrate severe inflow tract defects with pulmonary vein atresia (Supplementary Figs 13 and 14). Lineage tracing analyses revealed a statistically significant reduction of vascular and airway smooth muscle derived from *Isl1*⁺ and *Gli1*⁺ CPPs lacking *Smo* expression compared to controls (Fig. 4e–l).

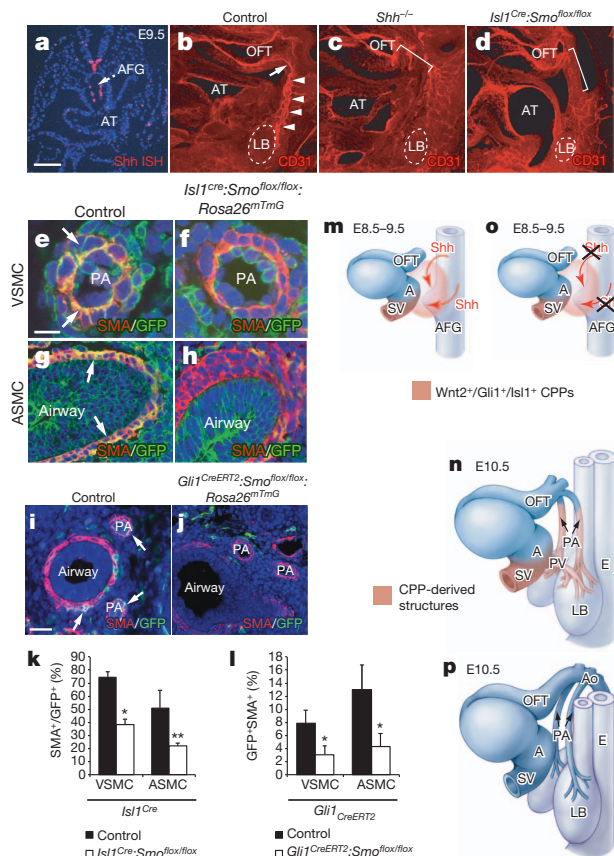


Figure 4 | Hedgehog signalling is required in CPPs to coordinate the vascular connection between the heart and lung. a–p, *Shh* expression at E9.5 in the anterior foregut endoderm by *in situ* hybridization (a). *Shh*^{-/-} mutants and *Isl1*^{Cre}:*Smo*^{flax/flax} mutants display a disrupted vascular plexus between the developing heart and lung at E10.5 (b–d, brackets; control, *Shh*^{+/+}). Inactivation of *Smo* within *Isl1*⁺ CPPs inhibits their contribution to airway and vascular smooth muscle at E13.5 (e–h, k; control, *Isl1*^{Cre}:*Smo*^{flax/+}; *Rosa26*^{mTmG}). Inactivation of *Smo* within *Gli1*⁺ CPPs inhibits their contribution to airway and vascular smooth muscle at E13.5 (i, j, l; control, *Gli1*^{CreERT2}:*Smo*^{flax/+}; *Rosa26*^{mTmG}). CPPs orchestrate lung and cardiac co-development and are regulated in turn by *Shh* expression from the anterior foregut endoderm (m–p). Ao, aorta; IVC, inferior vena cava; SVC, superior vena cava. **P* < 0.01, ***P* < 0.05. Error bars, s.d. Scale bars, 50 μ m (a), 20 μ m (e–h), 50 μ m (i and j).

We have used cell lineage tracing coupled with clonal analysis, to identify and characterize a novel population of mesoderm progenitors called CPPs that generate both cardiac inflow tract and pulmonary mesoderm cell lineages (Fig. 4m–p). Furthermore, we show that hedgehog signalling is required for the development of CPPs into the smooth muscle compartment in the lung and promotes the cardiopulmonary connection essential for terrestrial existence.

METHODS SUMMARY

The *Wnt2*^{CreERT2} mouse line was generated by insertion of the CreERT2 complementary DNA downstream of the ATG of the mouse *Wnt2* gene using homologous recombination in mouse embryonic stem cells as described previously¹⁵. Generation and genotyping of the *Gli1*^{CreERT2}, *Nkx2.5*^{Cre}, *Isl1*^{Cre}, *Rosa26*^{confetti}, *Smo*^{flax/flax}, *Rosa26*^{mTmG}, *Tie2*^{Cre}, *Sm22*^{Cre}, *VE-Cadherin*^{CreERT2}, *Shh*^{Cre}, *Ctnnb1*^{flax/flax}, *Gli1*^{LacZ/+} and *Isl1*^{MerCreMer} lines have been described previously^{4,5,9,11,16–21}.

Histological analysis of mouse embryos was performed as described²². A previously published protocol for immunostaining of thick tissue sections embedded in agarose was used²³.

Full Methods and any associated references are available in the online version of the paper.

Received 19 December 2012; accepted 7 June 2013.

Published online 21 July 2013.

- Peng, T. & Morrisey, E. E. Development of the pulmonary vasculature: current understanding and concepts for the future. *Pulm Circ* **3**, 176–178 (2013).
- Goss, A. M. *et al.* Wnt2/2b and β -catenin signaling are necessary and sufficient to specify lung progenitors in the foregut. *Dev. Cell* **17**, 290–298 (2009).
- Harris-Johnson, K. S., Domyan, E. T., Vezina, C. M. & Sun, X. β -Catenin promotes respiratory progenitor identity in mouse foregut. *Proc. Natl Acad. Sci. USA* **106**, 16287–16292 (2009).
- Cai, C. L. *et al.* *Isl1* identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. *Dev. Cell* **5**, 877–889 (2003).
- Moses, K. A., DeMayo, F., Braun, R. M., Reecy, J. L. & Schwartz, R. J. Embryonic expression of an *Nkx2-5*/Cre gene using ROSA26 reporter mice. *Genesis* **31**, 176–180 (2001).
- Goss, A. M. *et al.* Wnt2 signaling is necessary and sufficient to activate the airway smooth muscle program in the lung by regulating myocardin/Mrtf-B and Fgf10 expression. *Dev. Biol.* **356**, 541–552 (2011).
- Tian, Y. *et al.* Characterization and *in vivo* pharmacological rescue of a Wnt2-Gata6 pathway required for cardiac inflow tract development. *Dev. Cell* **18**, 275–287 (2010).
- Hoffmann, A. D., Peterson, M. A., Friedland-Little, J. M., Anderson, S. A. & Moskowitz, I. P. Sonic hedgehog is required in pulmonary endoderm for atrial septation. *Development* **136**, 1761–1770 (2009).
- Wang, Y. *et al.* Ephrin-B2 controls VEGF-induced angiogenesis and lymphangiogenesis. *Nature* **465**, 483–486 (2010).
- Red-Horse, K., Ueno, H., Weissman, I. L. & Krasnow, M. A. Coronary arteries form by developmental reprogramming of venous cells. *Nature* **464**, 549–553 (2010).
- Snippert, H. J. *et al.* Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* **143**, 134–144 (2010).
- Lavine, K. J., Long, F., Choi, K., Smith, C. & Ornitz, D. M. Hedgehog signaling to distinct cell types differentially regulates coronary artery and vein development. *Development* **135**, 3161–3171 (2008).
- White, A. C., Lavine, K. J. & Ornitz, D. M. FGF9 and SHH regulate mesenchymal Vegfa expression and development of the pulmonary capillary network. *Development* **134**, 3743–3752 (2007).
- Lin, L., Bu, L., Cai, C. L., Zhang, X. & Evans, S. *Isl1* is upstream of sonic hedgehog in a pathway required for cardiac morphogenesis. *Dev. Biol.* **295**, 756–763 (2006).
- Morrissey, E. E. *et al.* GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes Dev.* **12**, 3579–3590 (1998).
- Bai, L. Y. *et al.* Differential expression of Sonic hedgehog and *Gli1* in hematological malignancies. *Leukemia* **22**, 226–228 (2008).
- Harfe, B. D. *et al.* Evidence for an expansion-based temporal *Shh* gradient in specifying vertebrate digit identities. *Cell* **118**, 517–528 (2004).
- Jeong, J., Mao, J., Tenzen, T., Kottmann, A. H. & McMahon, A. P. Hedgehog signaling in the neural crest cells regulates the patterning and growth of facial primordia. *Genes Dev.* **18**, 937–951 (2004).
- Lepore, J. J. *et al.* High-efficiency somatic mutagenesis in smooth muscle cells and cardiac myocytes in SM22 α -Cre transgenic mice. *Genesis* **41**, 179–184 (2005).
- Long, F., Zhang, X. M., Karp, S., Yang, Y. & McMahon, A. P. Genetic manipulation of hedgehog signaling in the endochondral skeleton reveals a direct role in the regulation of chondrocyte proliferation. *Development* **128**, 5099–5108 (2001).
- Sun, Y. *et al.* *Isl1* is expressed in distinct cardiovascular lineages, including pacemaker and coronary vascular cells. *Dev. Biol.* **304**, 286–296 (2007).
- Shu, W., Jiang, Y. Q., Lu, M. M. & Morrisey, E. E. Wnt7b regulates mesenchymal proliferation and vascular development in the lung. *Development* **129**, 4831–4842 (2002).
- Snippert, H. J., Schepers, A. G., Delconte, G., Siersema, P. D. & Clevers, H. Slide preparation for single-cell-resolution imaging of fluorescent proteins in their three-dimensional near-native environment. *Nature Protocols* **6**, 1221–1228 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors appreciate the input of M. Kahn and J. Epstein in these studies. The authors are grateful to A. Stout for help in imaging. L. Guo provided assistance with figure illustrations. These studies were supported by funds from the National Institutes of Health (HL110942, HL100405, HL087825 to E.E.M. and HL117649 to S.M.E.) and the American Heart Association Jon DeHaan Myogenesis Center. T.P. is supported by T32 HL07586-23. C.J.B. is supported by P30 NS047101.

Author Contributions T.P. and E.E.M. designed the experiments. T.P., Y.T., C.J.B., M.M.L., R.S.K. and K.M.S. performed experiments. S.M.E. provided animal lines and expertise. T.P. and E.E.M. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.E.M. (emorrise@mail.med.upenn.edu).

METHODS

Animals. The *Wnt2^{CreERT2}* mouse line was generated by insertion of the CreERT2 cDNA downstream of the ATG of the mouse *Wnt2* gene using homologous recombination in mouse embryonic stem cells as described previously¹⁵. Schematic of targeting construct and Southern blot probes are shown in Supplementary Fig. 4. (PCR genotyping primers for *Wnt2^{CreERT2}* are: common forward primer: 5'-TGAGT CTCACCACTAGCCGCA-3'; *Wnt2^{CreERT2}* wild-type reverse: 5'-ACTGGGAATC AGCCAGGGAGGGT-3'; *Wnt2^{CreERT2}* mutant reverse: 5'-TCCAGGTATGCTCA GAAAACG-3'. Generation and genotyping of the *Gli1^{CreERT2}*, *Nkx2.5^{Cre}*, *Isl1^{Cre}*, *Rosa26^{confetti}*, *Smad^{flax/flax}*, *Rosa26^{mTmG}*, *Tie2^{Cre}*, *Sm22^{Cre}*, *VE-Cadherin^{CreERT2}*, *Shh^{Cre}*, *Ctnnb1^{flax/flax}*, *Gli1^{LacZ/+}* and *Isl1^{MerCreMer}* lines have been described previously^{4,5,9,11,16–21}. All animal experiments were carried out with approval of the University of Pennsylvania Institutional Animal Care and Use Committee (IACUC).

Histological analysis. At least three mouse embryos per genotype were collected at each of the time points (mentioned earlier) and fixed in 2% paraformaldehyde, dehydrated in a series of increasing ethanol concentration washes, embedded in paraffin and sectioned (except the experiments performed with the *Isl1^{MerCreMer}*; *Rosa26^{tdTomato}* embryos, which were performed with frozen sectioning as described previously²¹). Antibodies used are anti-smooth muscle actin (mouse anti-SMA 1:200; Abcam), CD31 (rat anti-CD31 1:500; BD Pharmingen), Von Willibrand Factor (rabbit anti-VWF 1:200; Sigma), MF20 (mouse anti-MF20 1:20; Abcam), Isl1 (mouse anti-Isl1 1:10; HybridomaBank), *Nkx2.5* (goat anti-*Nkx2.5* 1:10; Santa Cruz), Sarcomeric α -actinin (mouse anti-SAA 1:100; Sigma), NG2 (rabbit anti-NG2 1:100; Millipore) and GFP (goat anti-GFP 1:100; Abcam). LacZ staining of embryos was performed as described previously²². A previously published protocol for visualizing *Rosa26^{confetti}* expression in thick tissue sections embedded in agarose was used²³. *In situ* hybridization was performed as described previously²⁴. For whole-mount immunohistochemistry of embryos, a previously published protocol was employed²⁵. Embryos were fixed in 2% paraformaldehyde for 20 min, and washed with cold PBS three times for 10 min each. The washed embryos were then washed with blocking solution (0.4% Triton X-100, 5% serum in PBS) for 1 h on ice, and incubated with primary antibody (rat anti-CD31 1:500, BD Pharmingen; rabbit anti-GFP 1:1000, MBL; goat anti-GFP 1:500, Abcam; rabbit anti-SM22 α 1:250, Abcam) in blocking solution overnight. The next day, embryos were washed in 0.4% Triton X-100 in PBS on ice three times for 1 h each and incubated with secondary antibody overnight. The next day, the embryos were washed on ice three times for 1 h each. The stained embryos were then dehydrated in 100% methanol, and then BABB (1 part benzyl alcohol: 2 part benzyl benzoate) and mounted on a slide with Fastwell and sealed with cover slip. The slide was imaged on a Zeiss LSM 710 confocal microscope and analysed and reconfigured in ImageJ software.

Imaging of *Gli1^{CreERT2}*:*Rosa26^{confetti}* embryos with immunohistochemistry. A previously published protocol for immunostaining of thick tissue sections embedded in agarose was used²³. Embryos were fixed in 4% paraformaldehyde overnight

and washed with cold PBS four times for 30 min each. The embryos were then embedded in 4% agarose and sectioned on a vibratome to obtain 300- μ m thick slices. The individual slices embedded in agarose were then incubated in blocking solution (0.4% Triton X-100, 5% serum in PBS) for 1 h on ice, and incubated with primary antibody (rat anti-CD31 1:500, BD Pharmingen; rabbit anti-SM22 α 1:250, Abcam) in blocking solution overnight. The next day, slices were washed in 0.4% Triton X-100 in PBS on ice three times for 1 h each and incubated with secondary antibody overnight. The next day, the slices were washed on ice three times for 1 h each, and then mounted on a glass slide with Prolong Gold anti-fade reagent (Molecular Probes) and sealed with cover slip. The slides were imaged on a Zeiss LSM 710 confocal microscope and analysed in ImageJ software.

Clonal analysis. To assess for clonal relationships in CPPs, the *Wnt2^{CreERT2}*:*Rosa26^{mTmG}* were treated with tamoxifen at E8.5 at a concentration (0.025 mg g⁻¹ body weight) that was empirically determined to generate small clones or cluster of cells one to three in number at 24 h after treatment. A total of 59 embryos between E10.5 to E11.5 were determined to contain GFP clusters, and each was imaged per whole mount immunohistochemistry protocol with CD31/GFP double staining described above. Z-stacks were acquired for the entire volume capturing the heart and the lung to include all of the GFP⁺ cells. From 59 embryos, 22 clusters of GFP⁺ cells were identified that were spatially segregated with cellular distribution in the lung mesenchyme. Cells per cluster were counted through all Z-stacks and doubling time was calculated ($DT = (t - t_0) / \log_2((\log N - \log N_0))$). For clonal analysis of *Gli1^{CreERT2}*:*Rosa26^{confetti}* mice, 0.05 mg g⁻¹ body weight of tamoxifen was injected at E8.5 and tissue was collected at the indicated time points and sectioned at a thickness of 300 μ m and imaged per protocol described above with CD31 or SM22 α counterstaining. Z-stacks were acquired for the entire volume capturing the entire thickness of the section to include all fluorescently labelled cells. From 23 embryos, 15 clusters of single-coloured and spatially segregated cell clusters with distribution in the lung mesenchyme were identified for cellular localization, and 19 clusters for concurrent lineage analysis.

Cell counting. Sections included in cell count analysis were acquired using confocal microscopy. At least three embryos per genotype per time point were used and at least five sections were used per embryo at comparable levels. A minimum of 300 cells were counted per embryo. Cell counts were performed on ImageJ using the 'Cell Counter' plug-in and performed by two people blinded to the specimen genotype and condition. Results were averaged between the two cell counters and standard deviations were calculated per genotype/time point. One-tailed paired *t*-tests were used to determine the *P* value.

24. Morrissey, E. E., Ip, H. S., Lu, M. M. & Parmacek, M. S. GATA-6: a zinc finger transcription factor that is expressed in multiple cell lineages derived from lateral mesoderm. *Dev. Biol.* **177**, 309–322 (1996).
25. Yokomizo, T. *et al.* Whole-mount three-dimensional imaging of internally localized immunostained cells within mouse embryos. *Nature Protocols* **7**, 421–431 (2012).

Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing

Zhigang Xue^{1*}, Kevin Huang^{2*}, Chaochao Cai², Lingbo Cai³, Chun-yan Jiang³, Yun Feng¹, Zhenshan Liu¹, Qiao Zeng¹, Liming Cheng¹, Yi E. Sun¹, Jia-yin Liu³, Steve Horvath² & Guoping Fan²

Mammalian pre-implantation development is a complex process involving dramatic changes in the transcriptional architecture^{1–4}. We report here a comprehensive analysis of transcriptome dynamics from oocyte to morula in both human and mouse embryos, using single-cell RNA sequencing. Based on single-nucleotide variants in human blastomere messenger RNAs and paternal-specific single-nucleotide polymorphisms, we identify novel stage-specific mono-allelic expression patterns for a significant portion of polymorphic gene transcripts (25 to 53%). By weighted gene co-expression network analysis^{5,6}, we find that each developmental stage can be delineated concisely by a small number of functional modules of co-expressed genes. This result indicates a sequential order of transcriptional changes in pathways of cell cycle, gene regulation, translation and metabolism, acting in a step-wise fashion from cleavage to morula. Cross-species comparisons with mouse pre-implantation embryos reveal that the majority of human stage-specific modules (7 out of 9) are notably preserved, but developmental specificity and timing differ between human and mouse. Furthermore, we identify conserved key members (or hub genes) of the human and mouse networks. These genes represent novel candidates that are likely to be key in driving mammalian pre-implantation development. Together, the results provide a valuable resource to dissect gene regulatory mechanisms underlying progressive development of early mammalian embryos.

Mammalian pre-implantation development involves sequential decay of maternally stored RNAs in parallel with massive induction of new transcripts from the embryonic genome in a process called zygotic (or embryonic) genome activation (ZGA or EGA)^{7–9}. However, previous gene expression profiling of human pre-implantation has been limited by blastomere sample size and quantitation platforms^{10–13}. Recent advances in single-cell RNA sequencing (RNA-seq) technology^{14–17} have provided the unprecedented opportunity to study gene regulation in early human embryos under high resolution. We mapped approximately 700 million sequencing reads derived from 33 single cells of multiple stages, ranging from mature oocytes to 8-cell blastomeres. To control the quality of single-cell RNA-seq from technical errors, we focused on libraries (26 out of a total 33, or 79%) that share at least 80% similarity with other libraries derived from identical pre-implantation stages (Supplementary Table 1). In addition, we performed RNA-seq for three samples of individual morula embryos. The typical number of detectable genes (reads per kilobase per million (RPKM) > 0.1) ranged approximately from 8,500 to 12,000 genes in individual cells, demonstrating that our technique has greater sensitivity and coverage compared with previous microarray methods^{10–13}. Unsupervised hierarchical clustering and principal component analysis revealed that cells of different pre-implantation stages form distinct clusters (Fig. 1a, b). Furthermore, single-cell resolution revealed that blastomeres from the same 8-cell embryo are more similar to each other than blastomeres from a separate 8-cell embryo (Fig. 1a, b).

Intriguingly, cluster analyses showed that both pronuclear and zygote (1-cell stage) embryos clustered closely together but away from oocytes and cleavage embryos, indicating that the 1-cell stage exhibits a distinct transcriptome pattern. Comparisons between mature oocytes and 1-cell embryos identified 149 differentially expressed genes (false discovery rate (FDR) < 5%, >2-fold change), 79 and 70 of which were up- and downregulated in 1-cell embryos, respectively (Fig. 1c). Interestingly, approximately half of the upregulated genes remain highly expressed in 2- and 4-cell stages but are downregulated at the 8-cell stage, whereas the other half is further upregulated at the 8-cell stage (Fig. 1d). By performing similar single-cell RNA-seq in mouse pre-implantation embryos (Supplementary Table 1), we observed 520 transcripts that were upregulated in pronuclei compared to mature oocytes (Supplementary Fig. 1), confirming that both human and mouse species exhibit a conserved minor wave of ZGA before the major wave of ZGA (or EGA). These results indicated that our approach has overall greater

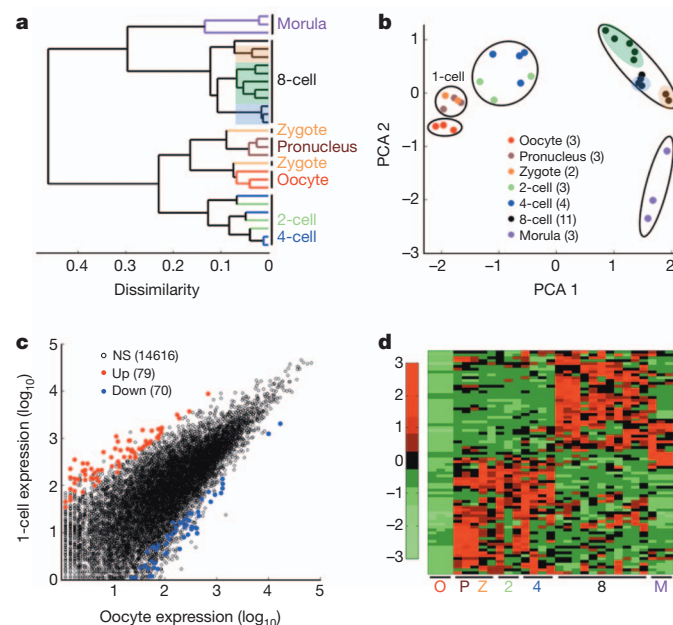


Figure 1 | High-resolution single-cell transcriptome analysis of human pre-implantation embryos. **a**, Unsupervised hierarchical clustering. **b**, Principal component analysis of single blastomere expression patterns for seven stages of oocytes and pre-implantation embryos. Sister cells from the same 8-cell embryo are highlighted together. The number of samples for each stage are indicated in the PCA legend in parenthesis as well as in Supplementary Table 1. **c**, Scatterplot showing the number of activated (red) and reduced (blue) genes in 1-cell embryos ($n = 5$) compared to oocytes ($n = 3$). **d**, Heatmap showing relative expression patterns of 1-cell activated genes ($n = 79$) across all pre-implantation stages. NS, not significant.

¹Translational Center for Stem Cell Research, Tongji Hospital, Department of Regenerative Medicine, Tongji University School of Medicine, Shanghai 200065, China. ²Department of Human Genetics, David Geffen School of Medicine, UCLA, Los Angeles, California 90095, USA. ³State Key Laboratory of Reproductive Medicine, Center of Clinical Reproductive Medicine, First Affiliated Hospital, Nanjing Medical University, Nanjing 210029, China.

*These authors contributed equally to this work.

quantitation for identifying gene expression changes between different stages of pre-implantation development.

Single-cell RNA-seq enabled base-resolution scrutiny without confounding effects from cell population heterogeneity. In this unique study, all embryos were derived from intra-cytoplasmic sperm injection of different egg donors with the same sperm donor. Based on the paternal genotype as assayed by exome sequencing of the sperm donor's blood sample, we were able to follow the paternal genome in embryos through single-nucleotide polymorphism (SNP) analysis^{18–20}. We identified paternally or maternally expressed genes either by inferring phased paternal haplotype information^{21,22} or by leveraging sites which are paternally homozygous in exome sequencing, but heterozygous in embryo transcripts (Fig. 2a and Supplementary Table 2). In total, we determined the parent-of-origin expression for approximately 850 to 1,400 gene transcripts (or 15–20% of all detected gene transcripts per stage). For further validation of allele-specific expression in early embryos, we analysed the status of imprinting genes that exhibit known parent-of-origin expression patterns (see Supplementary Text and Supplementary Figs 2 and 3). Unexpectedly, we found that 53% of 8-cell embryo transcripts and 23% of morula transcripts exhibit monoallelic maternal expression patterns (Fig. 2b), even though major EGA in human early embryos already occurred during the transition of 4- to 8-cell stage (Supplementary Fig. 4). For example, within a single haplotype region, we observed simultaneous maternal transcript activation and degradation in the *ASB6* locus and *C9orf78* locus, respectively (Fig. 2c). In a different scenario, the paternal genotype is homozygous in three consecutive SNPs (rs3829009, rs6990278 and rs8537) at the cell-cycle regulator *CDCA2* gene locus. In this example, the phased paternal SNP pattern is not detected in 2- and 4-cell embryos but seen in 8-cell embryos, indicating that the detected transcripts in 2- and 4-cell embryos must be of maternal origin and also that *CDCA2* undergoes paternal activation at the 8-cell stage (Fig. 2d). However, paternal activation at this locus appears transient as the paternal allele could not be detected in the morula stage. In fact, we observed a moderate correlation between expression of the paternal allele and overall transcript expression ($r = 0.53$, $P = 0.06$), suggesting that the paternal allele is regulated dynamically.

Notably, the rs3829009 SNP in *CDCA2* is a missense variant that results in a benign Arg to Ser amino acid change²³ (Fig. 2d). As RNA-seq produces base-resolution information for codons, we further examined the global prevalence of potentially deleterious transcript expression. Using multiple single-cells from the same 8-cell embryo, we consistently identified 1,225 homozygous SNP variants, 3.5% of which are predicted as damaging non-synonymous variants (Supplementary Fig. 5). This result suggests that single-cell RNA-seq would be a powerful approach for genome-scale screening of potentially deleterious variants in embryos.

Thus far the analyses focused on individual gene-transcript changes during each transitional stage, but did not reveal the crucial shift of gene networks in pre-implantation development. To understand the co-expression relationships between genes at a systems level, we performed weighted gene co-expression network analysis (WGCNA)^{5,6}. This unsupervised and unbiased analysis identified distinct co-expression modules corresponding to clusters of correlated transcripts (Fig. 3a). Notably, 9 out of 25 co-expression modules showed stage-specific expression values, that is, these modules are comprised of genes that tend to be overexpressed in a single developmental stage ($r > 0.7$, $P < 10^{-3}$, Supplementary Fig. 6).

These stage-specific modules probably represent core gene networks operating in each transitional stage. In total, 8,301 genes were part of human oocyte to morula stage-specific modules, revealing a step-wise requirement of new transcripts that are involved in gene transcription (4-cell), post-transcriptional RNA processing (8-cell), and then protein translation and cell energetics (morula) (Fig. 3b). As expected, the oocyte module undergoes gradual degradation over the course of development, whereas modules in the 4- to 8-cell transition show sharp degradation (threefold) and activation (fourfold) (Fig. 3c). Collectively, our systems analysis revealed that the transcriptional organization for human pre-implantation development can be summarized using a small number of stage-specific modules with well-defined function.

Importantly, these modules are highly robust and reproducible as can be seen from our module preservation analysis²⁴ (Fig. 4a). This module preservation analysis uses a permutation test to define a test

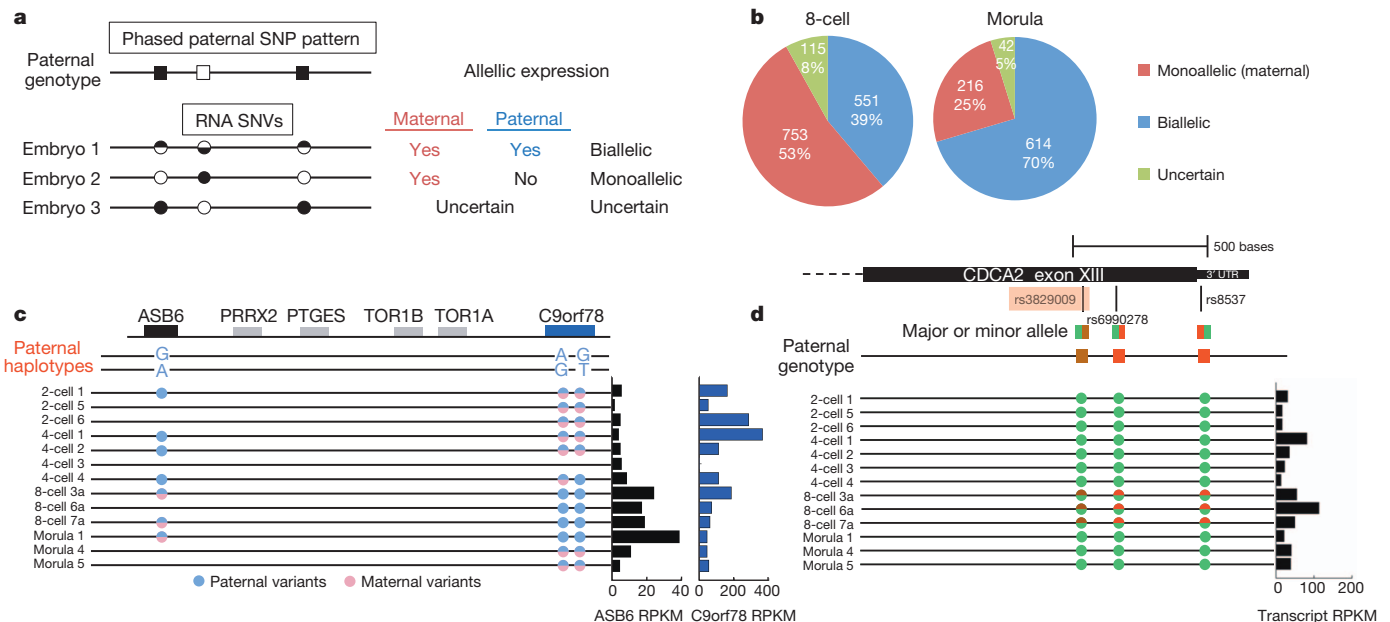


Figure 2 | Tracing parent-of-origin allelic RNA transcripts through SNP analysis in pre-implantation embryos. **a**, Schematic for deducing maternal and paternal transcript origin from the presence or absence of homozygous paternal alleles. Embryo number 3 has unclear contribution from the maternal allele. **b**, Pie charts showing the number and per cent of all gene transcripts exhibiting single-nucleotide variants (SNVs) and their assignments into one of

three categories as illustrated in **a**. **c**, Example of maternal activation and degradation from a deduced partial haplotype. **d**, Example of paternal activation by comparing embryos that showed biallelic expression in one stage but monoallelic (maternal) expression in a previous stage. rs3829009 is highlighted because the minor allele is an Arg884Ser missense variant. SNPs are adenine (green), guanine (red) and thymine (brown).

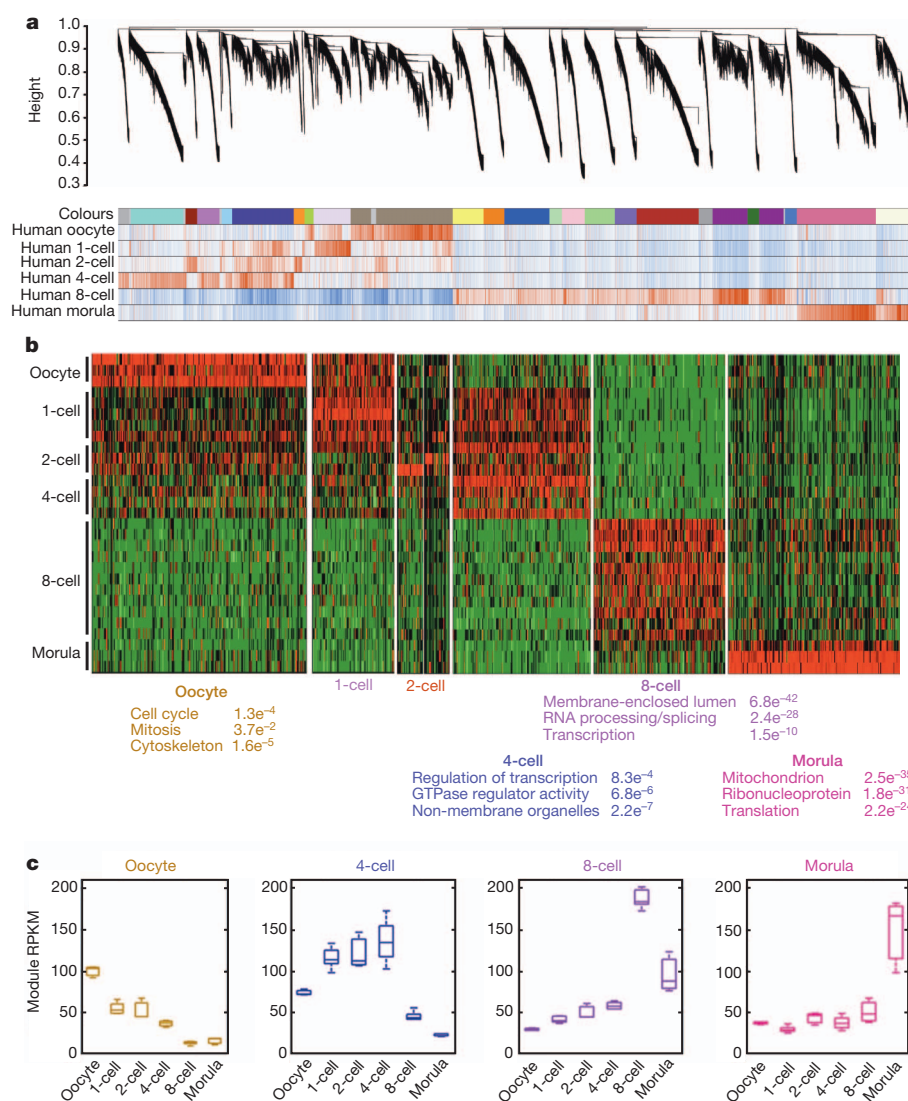


Figure 3 | Network analysis of human pre-implantation development.

a, Hierarchical cluster tree showing co-expression modules identified using WGCNA. Modules correspond to branches and are labelled by colours as indicated by the first colour band underneath the tree. Remaining colour bands reveal highly correlated (red) or anti-correlated (blue) transcripts for the particular stage. **b**, Heatmap showing relative expression of 7,313 genes in

7 representative stage-specific modules across all samples. As each developmental window only has one or two highly correlated modules, the modules were assigned biological names. Top three representative gene ontology terms and their associated functional enrichment *P* values are shown below. **c**, Boxplots showing the distribution of module expression (mean RPKM of all genes within a given module) for different cell types.

statistic, Z_{summary} , that summarizes the evidence that the network topology of the module is preserved in an independent data set (see Methods). Seven out of nine modules were detected with highly significant preservation scores ($Z_{\text{summary}} > 10$) in two publicly available human preimplantation data sets (but with substantially less RNA transcripts). Prior to the direct study of transcription changes during human preimplantation process, mouse models provided crucial insights into the pre-implantation gene regulatory network^{2–4}. However, it seems that only one study (involving far fewer transcripts) performed a cross-species comparison in the context of a pluripotency network¹¹. Therefore, we performed single-cell RNA-seq in mouse blastomeres from oocyte to morula embryos for a direct cross-species comparative analysis (Supplementary Table 1). Cross-species module preservation analysis showed that 7 out of 9 human stage-specific modules were at least moderately preserved ($Z_{\text{summary}} > 5$) (Fig. 4a and Supplementary Table 1). We further validated the cross-species preservation using other mouse pre-implantation data sets^{11,25}, which confirmed most human stage-specific modules are preserved in mouse (Fig. 4a). As expected, intra-species module preservation was more significant ($Z_{\text{summary}} > 10$) than inter-species preservation ($Z_{\text{summary}} > 5$, Fig. 4a).

Thus, these preserved gene networks represent a strongly conserved transcriptional architecture of key developmental programs.

By applying WGCNA independently to our mouse data set, we found that mouse development also involves stage-specific co-expression modules (Supplementary Fig. 7). Gene ontology analyses showed that mouse stage-specific modules share many functional similarities to their human counterparts, including the conserved sequential activation of functional enrichment changes (Supplementary Fig. 8). As expected, overlaps between human and mouse stage-specific modules were highly significant ($P < 10^{-4}$, Fig. 4b, see also Supplementary Fig. 9), and the functional enrichment of overlapped stage-specific genes mostly reflected the overall network function (Fig. 4c). However, when we examined overlap of pre-major ZGA in human and mouse, we found that most genes were enriched for protein transport and GTPase signalling, whereas their respective modules as a whole were more enriched for cell-cycle genes (Fig. 4c and Supplementary Fig. 8).

Notably, stage-specific modules in human and mouse overlapped across multiple stages. For example, the mouse oocyte and 1-cell module genes overlapped significantly with genes that were specific to human oocyte, 1-cell and 4-cell stages ($P < 10^{-6}$, Fig. 4b). This result suggested

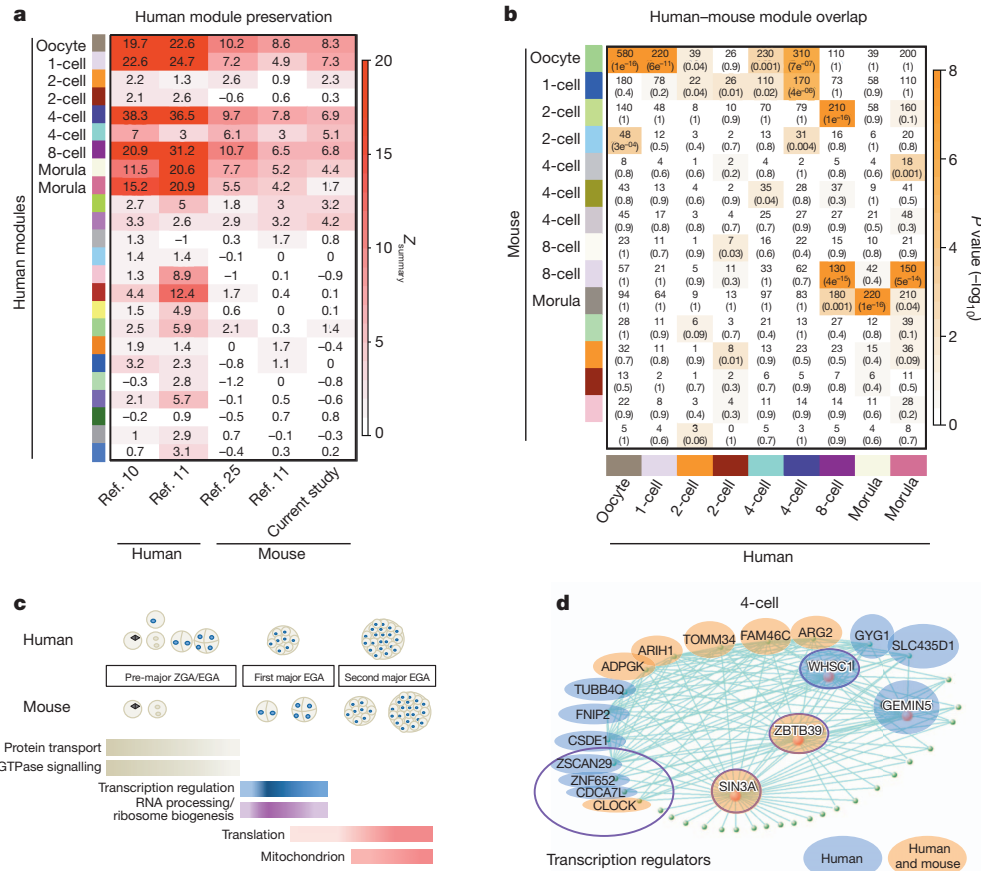


Figure 4 | Stage-specific gene activation is preserved in human and mouse pre-implantation development. **a**, Heatmap of human module preservation scores in available independent data sets labelled on the x axis. **b**, Heatmap showing the significance of gene overlaps between independently constructed human and mouse modules. The x axis shows only human stage-specific modules ($n = 9$) and the y axis shows all mouse modules ($n = 15$). Each cell contains the number of intersecting genes and P value of the intersection. Colour legend represents $-\log_{10} P$ value based on the hypergeometric test.

that mouse pre-major ZGA genes are spread over the longer gestational pre-major EGA window in humans. Likewise, post-major ZGA modules in mouse were found to have significant overlap and spread throughout all post-major EGA human stages (Fig. 4c). Although there is divergence in the timing of the major ZGA between human and mouse, these two species re-converge in both timing and function at the 8-cell to morula transition, during which many mitochondrion-related transcripts are upregulated (Fig. 4c and Supplementary Fig. 8). Collectively, our results show that human and mouse share many core transcriptional programs in early development, but diverge in their stage-specificity and timing, probably reflecting species-specific differences in human and mouse gestational periods.

Using the WGCNA measure of intramodular gene connectivity (kME), we identified 491 intramodular hub genes across all stage-specific modules ($kME > 0.9$, $P < 10^{-22}$). Intramodular hub genes are centrally located in their respective modules and may thus be critical components within the network. Remarkably, 337 (69%, $P < 10^{-16}$) hub genes in stage-specific modules can be validated (that is, independently identified as a hub gene in a separate data set), demonstrating that pre-implantation hub genes are highly reproducible (Supplementary Table 3). For example, KPNA7 ($kME = 0.89$, $P < 10^{-22}$) is consistently identified as a hub gene in multiple human and mouse pre-major EGA networks ($kME > 0.91$). Notably, approximately half of Kpna7-deficient mouse embryos fail to develop to blastocysts, whereas the other half show delayed developmental progress²⁶. Furthermore, SIN3A, a transcriptional co-repressor,

is a 4-cell intramodular hub gene that was validated in every independent data set (Fig. 4d and Supplementary Table 3). Indeed, analysis of upstream regulators of human 8-cell module genes revealed overrepresentation in MYC, MAX and MXI1 targets, which are absent in the 4-cell stage modules. This finding is consistent with the previously characterized activation of Myc-enriched genes upon Sin3a depletion²⁷. Together, these results demonstrate that some intramodular hubs are probably key players in pre-implantation development.

In summary, we have demonstrated that single-cell RNA sequencing has markedly improved transcriptome quantification of rare human pre-implantation embryo samples at both individual transcript and system levels. The findings extend our knowledge of the transcriptional architecture, sequential order of gene activation, and genetic programming for early human embryogenesis. Our cross-species systems analysis demonstrates that the human pre-implantation transcriptional organization is highly preserved, highlighting an evolutionarily conserved molecular process including key genes that drive mammalian pre-implantation development. It remains to be resolved whether the differences in a portion of co-expressed modules can account for species-specific function, such as seen previously in systems comparisons of human and mouse brain²⁸. Furthermore, we expect that single-cell RNA-seq can also quantitatively delineate the structures, isoforms, and allele-specific expression patterns of both coding genes and non-coding regulatory RNAs^{8,13,25}. Moreover, compared to exome or genomic sequencing, RNA-seq analysis has the advantage of quantitatively

revealing expression defects that are due to genetic and/or epigenetic alternations in gene regulatory domains that are beyond the detection of DNA sequencing, such as in the case of imprinting disorders. Thus, single-cell RNA-seq of a single blastomere could be a promising approach for pre-implantation genetic diagnosis in the near future.

METHODS SUMMARY

Oocytes and early embryos were obtained from the Center for Clinical Reproductive Medicine at the Jiangsu People's Hospital, Nanjing, China, with patients' written informed consent and institutional approval. Single-cell isolation and RNA-seq experiments were carried out as described¹⁴ with the use of the Illumina HiSeq2000 instrument, according to the manufacturer's specifications. Detailed information on ethical conduct, sample collection, and data analyses are available in Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 10 July 2012; accepted 10 June 2013.

Published online 28 July; corrected online 28 August 2013 (see full-text HTML version for details).

- Niakan, K. K., Han, J., Pedersen, R. A., Simon, C. & Pera, R. A. Human pre-implantation embryo development. *Development* **139**, 829–841 (2012).
- Hamatani, T., Carter, M. G., Sharov, A. A. & Ko, M. S. Dynamics of global gene expression changes during mouse preimplantation development. *Dev. Cell* **6**, 117–131 (2004).
- Wang, Q. T. *et al.* A genome-wide study of gene activity reveals developmental signaling pathways in the preimplantation mouse embryo. *Dev. Cell* **6**, 133–144 (2004).
- Zeng, F., Baldwin, D. A. & Schultz, R. M. Transcript profiling during preimplantation mouse development. *Dev. Biol.* **272**, 483–496 (2004).
- Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
- Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* <http://dx.doi.org/10.2202/1544-6115.1128> (2005).
- Walser, C. B. & Lipshitz, H. D. Transcript clearance during the maternal-to-zygotic transition. *Curr. Opin. Genet. Dev.* **21**, 431–443 (2011).
- Schier, A. F. The maternal-zygotic transition: death and birth of RNAs. *Science* **316**, 406–407 (2007).
- Schultz, R. M. The molecular foundations of the maternal to zygotic transition in the preimplantation embryo. *Hum. Reprod. Update* **8**, 323–331 (2002).
- Vassena, R. *et al.* Waves of early transcriptional activation and pluripotency program initiation during human preimplantation development. *Development* **138**, 3699–3709 (2011).
- Xie, D. *et al.* Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species. *Genome Res.* **20**, 804–815 (2010).
- Zhang, P. *et al.* Transcriptome profiling of human pre-implantation development. *PLoS ONE* **4**, e7844 (2009).
- Dobson, A. T. *et al.* The unique transcriptome through day 3 of human preimplantation development. *Hum. Mol. Genet.* **13**, 1461–1470 (2004).
- Tang, F. *et al.* RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nature Protocols* **5**, 516–535 (2010).
- Islam, S. *et al.* Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* **21**, 1160–1167 (2011).
- Ramskold, D. *et al.* Full-length mRNA-seq from single-cell levels of RNA and individual circulating tumor cells. *Nature Biotech.* **30**, 777–782 (2012).
- Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-seq by multiplexed linear amplification. *Cell Rep.* **2**, 666–673 (2012).
- The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
- McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
- Li, Y. *et al.* Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nature Genet.* **42**, 969–972 (2010).
- Gabriel, S. B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229 (2002).
- Altshuler, D. M. *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
- Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols* **4**, 1073–1081 (2009).
- Langfelder, P., Luo, R., Oldham, M. C. & Horvath, S. Is my network module preserved and reproducible? *PLoS Comput. Biol.* **7**, e1001057 (2011).
- Tang, F. *et al.* Deterministic and stochastic allele specific gene expression in single mouse blastomeres. *PLoS ONE* **6**, e21208 (2011).
- Hu, J. *et al.* Novel importin- α family member Kpna7 is required for normal fertility and fecundity in the mouse. *J. Biol. Chem.* **285**, 33113–33122 (2010).
- Dannenberg, J. H. *et al.* mSin3A corepressor regulates diverse transcriptional networks governing normal and neoplastic growth and survival. *Genes Dev.* **19**, 1581–1595 (2005).
- Miller, J. A., Horvath, S. & Geschwind, D. H. Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. *Proc. Natl Acad. Sci. USA* **107**, 12698–12703 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank many of our colleagues for invaluable discussions and comments on our study, in particular, P. Pajukanta at UCLA for very helpful suggestions on SNP analyses and D. Geschwind for critically reading the manuscript. This work was supported by 973 Grant Programs 2012CB966300, 2011CB966204 and 2011CB965102 from the Ministry of Science and Technology in China; the International Science and Technology Cooperation Program of China (no. 2011DFB30010); and the National Natural Science Foundation of China (81271258).

Author Contributions Z.X., K.H., J.L. and G.F. designed the study. Z.X., L.C., C.J., Y.F., Z.L., Q.Z., L.C. and Y.E.S. carried out experiments or contributed critical reagents and protocols. K.H., C.C. and S.H. analysed the data and performed statistical analyses. K.H. and G.F. wrote the manuscript in discussion with all the authors. All the authors read and approved the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.F. (gfan@mednet.ucla.edu), J.L. (jyliu_nj@126.com) or Z.X. (xuezhigang75@gmail.com). All sequencing data generated for this work have been deposited in the NCBI Gene Expression Omnibus (GEO) under accession number GSE441183.

METHODS

Ethics statement. Oocytes were obtained from female patients (between ages 26 and 35) at the Center for Clinical Reproductive Medicine at the Jiangsu People's Hospital, Nanjing, China, with written informed consent and institutional approval. Sperm was obtained from an anonymous healthy donor at an *in vitro* fertilization sperm bank in Changsha, China, with informed consent. This study was approved by the Institutional Review Board (IRB) on Human Subject Research and Ethics Committee in the First Affiliated Hospital to Nanjing Medical University, China. None of the donors received any financial re-imbursement. The identity of the subjects was anonymous to all of the scientists in this project from the beginning. Investigators at UCLA were only involved in data analysis and manuscript writing. According to the UCLA IRB review of their involvement in the study, UCLA investigators were granted exemption approval (IRB no. 12-001361).

Human embryo collection and culture. Vitified mature oocytes were warmed with a thawing kit before RNA isolation (Jieying Laboratory). All thawing steps were performed at room temperature (23–25 °C). Oocyte survival was evaluated based on integrity of the oocyte membrane and the zona pellucida. Fertilization is achieved through intracytoplasmic sperm injection (ICSI) to synchronize the development time for each batch of viable oocytes for this study. Embryos were cultured in Cleavage Medium (SAGE) in a low-oxygen humidified atmosphere containing 5% CO₂, 5% O₂ and 90% N₂ at 37 °C. Stages of human embryo development were assessed through microscopy and collected separately.

Mice embryos collection. For timed pregnancy, PMSG (7-10UI) was intraperitoneally injected into C57BL/6 female mice aged 4 to 6 weeks. Next, hCG (7-10UI) was intraperitoneally injected after 46 to 48 h. Pregnant mice were euthanized at various time points to obtain embryos as follows: metaphase II oocyte (12 h), zygote (24 to 26 h), 2-cell (30 to 32 h), 4-cell embryo (day 2), 8-cell embryo (day 3), and morula (day 4). Animal experiments were reviewed and approved by the Animal Experiment Administration Committee of Tongji University School of Medicine, China.

RNA isolation and library construction. Both human and mouse blastomeres were prepared using identical protocols. Single blastomeres were isolated by removing the zona pellucida using acidic tyrode solution (Sigma, catalogue no. T1788), then separated by gentle mouth pipetting in a calcium-free medium. Single cells were washed twice with 1 × PBS containing 0.1% BSA before placing in lysis buffer. RNA was isolated from single cells or single morula embryos and amplified as described previously¹⁴. Library construction was performed following Illumina manufacturer suggestions. Libraries were sequenced on the Illumina HiSeq2000 platform and sequencing reads that contained polyA, low quality, and adapters were pre-filtered before mapping. Filtered reads were mapped to the hg19 genome and mm9 genome using default parameters from BWA aligner²⁹, and reads that failed to map to the genome were re-mapped to their respective mRNA sequences to capture reads that span exons.

Transcriptional profiling. In both human and mouse cases, data normalization was performed by transforming uniquely mapped transcript reads to RPKM³⁰. Genes with low expression in all stages (average RPKM < 0.5) were filtered out, followed by quantile normalization. For differential expression, we compared every time point to its previous time point using default parameters in DESeq using normalized read counts. Genes were called differentially expressed if they exhibited a Benjamini and Hochberg-adjusted *P* value (FDR) < 5% and a mean fold change of > 2.

Paternal exon sequencing. The paternal genomic DNA was extracted from 10 ml peripheral blood of the sperm donor (serial number D5422) using QIAamp DNA Blood Mini Kit. Then 1 µg of the genomic DNA was fragmented through sonication. Exome enrichment was performed using the Agilent SureSelect Human All Exon Kit (50 Mb). Sequencing was performed on the Illumina HiSeq2000 platform according to the manufacturer's instructions and reads were mapped to the hg19 using default parameters for the BWA aligner²⁹.

SNP analysis. SNP calling was performed using the GATK software using default parameters¹⁶. The SNP database we used for reference was from the 1000 Genomes Project which contains SNP variants from the Chinese population (CHB). Only SNP sites with coverage of 10 reads or more were considered. SNP annotation was performed using Annovar³¹ and SNP variant effect on protein functional change was predicted using SIFT²³. Only predictions with high confidence (SIFT score < 0.05) were considered.

Paternal and maternal gene calling. We performed paternal and maternal gene calling in two ways. In the first, we deduced the parental haplotypes and used linkage information to determine the parent-of-origin. We used heterozygous SNPs from paternal genotype (first allele ≥ 5 reads, second allele ≥ 3 reads, both quality values > 20) and identified at least two embryos which contain paternal SNPs (both alleles ≥ 5 reads, both quality values > 20). We asked whether the linkage between two adjacent loci are supported by correct links in at least two embryos, and whether the correct links are greater than the wrong links at each loci (for example,

the number of correct links > 3 × wrong links). These results are in Supplementary Table 2.

In the second approach, we used homozygous SNPs from the paternal genotype so we could trace exactly which allele the embryo must carry. If embryos are heterozygous at this site, we infer that the alternative allele is maternally derived. In addition, if there is absence of the paternal allele and expression of an alternative allele, these transcripts are also maternally derived. We have provided a schematic for these scenarios in Fig. 2a.

Weighted gene co-expression network analysis. Both human and mouse data sets were independently constructed using the following method. A signed weighted correlation network was constructed by first creating a matrix of pairwise correlations between all pairs of genes across the measured samples⁶. Next, the adjacency matrix was constructed by raising the co-expression measure, $0.5 + 0.5 \times \text{correlation matrix}$, to the power $\beta = 12$. The power of 12, which is the default value, is interpreted as a soft-threshold of the correlation matrix. Based on the resulting adjacency matrix, we calculated the topological overlap, which is a robust and biologically meaningful measure of network interconnectedness³² (that is, the strength of two genes' co-expression relationship with respect to all other genes in the network). Genes with highly similar co-expression relationships were grouped together by performing average linkage hierarchical clustering on the topological overlap. We used the Dynamic Hybrid Tree Cut algorithm³³ to cut the hierarchical clustering tree, and defined modules as branches from the tree cutting. We summarized the expression profile of each module by representing it as the first principal component (referred to as module eigengene). Modules whose eigengenes were highly correlated (correlation above 0.7) were merged.

Module preservation statistics. To evaluate the human modules in mouse developmental data²⁴, we mapped human genes to mouse genes (orthologous genes) as annotated from the Mouse Genome Informatics (MGI) database³⁴. An advantage of WGCNA is that it provides powerful module preservation statistics which assess whether the density (how tight interconnections among genes in a module are) and connectivity patterns of individual modules (for example, intramodular hub gene status) are preserved between two data sets. To assess the preservation of human modules (reference network) in the test network (either human or mouse), we used the R function 'modulePreservation' in the WGCNA R package, as this quantitative measure of module preservation enables rigorous argument that a module is not preserved²⁴. By averaging the several preservation statistics generated through many permutations of the original data, a Z_{summary} value is calculated, which summarizes the evidence that a module is preserved and indicative of module robustness and reproducibility. In general, modules with Z_{summary} scores > 10 are interpreted as strongly preserved (that is, densely connected, distinct, and reproducible modules), Z_{summary} scores between 2 and 10 are weak to moderately preserved, and Z_{summary} scores < 2 are not preserved.

Identification and visualization of hub genes. Module eigengenes lead to a natural measure of module membership (also known as module eigengene based connectivity kME). Specifically, an approximate measure of module membership for gene i with respect to module q is defined as follows $MM^q(i) = \text{cor}(x(i), E^q)$, where $x(i)$ is the expression profile of gene i and E^q is the eigengene of module q . As we use signed networks here, we expect that module genes have significant positive module membership values. The advantage of using a correlation to quantify module membership is that this measure is naturally scaled to lie in the interval $[-1, 1]$ and a corresponding statistical significance measure (*P* value) can be easily computed. Genes with highest module membership values are referred to as intramodular hub genes (for example, $kME > 0.9$, $P < 10^{-22}$). Intramodular hub genes are centrally located inside the module and represent the expression profiles of the entire module³⁵. We used VisANT³⁶ to visualize the top 150 gene connections (based on topological overlap) among the top 100 hub genes (that is, genes with the highest kME).

Hub gene validation. We used WGCNA to independently construct a network in published data sets, and generated an independent list of hub genes ($kME > 0.9$) for each data set. For human–human comparisons and mouse–mouse comparisons, we determined the overlap of hub genes from the same developmental stage. However, for human–mouse module overlaps, we used all modules with significant gene overlap ($P < 10^{-4}$, see Fig. 4b) to compute intramodular hub gene overlap. For example, we overlap hub genes found in the human 4-cell module to both mouse oocyte and 1-cell networks since both these networks have significant overlaps with the 4-cell module.

Enrichment of upstream regulators. The Ingenuity Upstream Regulator Analytic was used to determine enrichment of upstream regulators for all human stage-specific modules. Regulators that were not transcriptional regulators were not considered.

Gene ontology analysis. Functional annotation was performed with the Database for Annotation, Visualization and Integrated Discovery (DAVID) Bioinformatics Resource³⁷. Gene ontology terms shown in this study summarized all similar

sub-terms into an overarching term, and Benjamini-Hochberg adjusted P values are shown for the representative term.

29. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
30. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**, 621–628 (2008).
31. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
32. Yip, A. M. & Horvath, S. Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinformatics* **8**, 22 (2007).
33. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719–720 (2008).
34. Eppig, J. T., Blake, J. A., Bult, C. J., Kadin, J. A. & Richardson, J. E. The Mouse Genome Database (MGD): comprehensive resource for genetics and genomics of the laboratory mouse. *Nucleic Acids Res.* **40**, D881–D886 (2012).
35. Horvath, S. & Dong, J. Geometric interpretation of gene coexpression network analysis. *PLOS Comput. Biol.* **4**, e1000117 (2008).
36. Hu, Z., Mellor, J., Wu, J. & DeLisi, C. VisANT: an online visualization and analysis tool for biological interaction data. *BMC Bioinformatics* **5**, 17 (2004).
37. Huang, D. W., Sherman, B. T. & Lempicki, R. A., Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc.* **4**, 44–57 (2009).

lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs

Liuqing Yang^{1,2*}, Chunru Lin^{1,2*}, Chunyu Jin¹, Joy C. Yang³, Bogdan Tanasa^{1,4}, Wenbo Li¹, Daria Merkurjev^{1,5}, Kenneth A. Ohgi¹, Da Meng⁶, Jie Zhang¹, Christopher P. Evans³ & Michael G. Rosenfeld¹

Although recent studies have indicated roles of long non-coding RNAs (lncRNAs) in physiological aspects of cell-type determination and tissue homeostasis¹, their potential involvement in regulated gene transcription programs remains rather poorly understood. The androgen receptor regulates a large repertoire of genes central to the identity and behaviour of prostate cancer cells², and functions in a ligand-independent fashion in many prostate cancers when they become hormone refractory after initial androgen deprivation therapy³. Here we report that two lncRNAs highly overexpressed in aggressive prostate cancer, PRNCR1 (also known as PCAT8) and PCGEM1, bind successively to the androgen receptor and strongly enhance both ligand-dependent and ligand-independent androgen-receptor-mediated gene activation programs and proliferation in prostate cancer cells. Binding of PRNCR1 to the carboxy-terminally acetylated androgen receptor on enhancers and its association with DOT1L appear to be required for recruitment of the second lncRNA, PCGEM1, to the androgen receptor amino terminus that is methylated by DOT1L. Unexpectedly, recognition of specific protein marks by PCGEM1-recruited pygopus 2 PHD domain enhances selective looping of androgen-receptor-bound enhancers to target gene promoters in these cells. In 'resistant' prostate cancer cells, these overexpressed lncRNAs can interact with, and are required for, the robust activation of both truncated and full-length androgen receptor, causing ligand-independent activation of the androgen receptor transcriptional program and cell proliferation. Conditionally expressed short hairpin RNA targeting these lncRNAs in castration-resistant prostate cancer cell lines strongly suppressed tumour xenograft growth *in vivo*. Together, these results indicate that these overexpressed lncRNAs can potentially serve as a required component of castration-resistance in prostatic tumours.

One of the overexpressed lncRNAs in prostate cancer, PCGEM1, is tissue-specific and correlated with high-risk prostate cancer patients, including African-American men⁴, whereas a second highly expressed lncRNA, PRNCR1 (PCAT8), is pervasively transcribed from the 8q24 'gene desert' region in strong association with susceptibility of prostate cancer⁵. Paired benign prostatic hyperplasia and aggressive tumour specimens (Gleason scores 3 + 3) derived from three individual prostate cancer patients exhibited >100-fold upregulation of PRNCR1 and PCGEM1 expression (Supplementary Fig. 1a, b). Native RNA immunoprecipitation (RIP) experiments in paired prostate tumour and benign prostatic hyperplasia tissues (Gleason scores 2 + 3 to 4 + 3), revealed a specific association of full-length androgen receptor (AR) with both PRNCR1 and PCGEM1 in prostate tumour tissues (Fig. 1a, b; Supplementary Fig. 1a, c) compared to minimal interactions with glucocorticoid receptor (Supplementary Fig. 1c and data not shown). In prostate cancer LNCaP cells, dihydrotestosterone (DHT) treatment induced association of AR with both PRNCR1 and PCGEM1 (Fig. 1c), but not with NEAT2 (also known as MALAT1) (Fig. 1c). Antisense

oligonucleotide (ASO)-based knockdown of PRNCR1 abolished both AR-PRNCR1 and AR-PCGEM1 interactions, whereas knockdown of PCGEM1 did not affect the AR-PRNCR1 interaction (Fig. 1d; Supplementary Fig. 2a), indicating the PRNCR1-dependent recruitment of PCGEM1.

Knockdown of either PRNCR1 or PCGEM1 resulted in a significant decrease in DHT-induced activation of canonical AR target genes while not affecting AR levels (Supplementary Fig. 2a–c). Global run-on sequencing (GRO-Seq) confirmed that knockdown of either PRNCR1 or PCGEM1 significantly decreased the induction of 617 DHT-upregulated genes ($n = 617$, edgeR false discovery rate (FDR) < 0.01, and read density (RD) > 2) with AR-bound enhancers within 200 kilobases (kb) (Fig. 1e), but had no effect on DHT-unresponsive genes located >200 kb away from AR-bound enhancers (Supplementary Fig. 2d), verified by randomly extracting sets of 1,000 genes (data not shown).

Using chromatin isolation by RNA purification (ChIRP)⁶ with biotin-labelled DNA probes (40–60 nucleotides) tiling PRNCR1 and PCGEM1 RNAs (Supplementary Fig. 3a–c), we identified 2,142 high-confidence PCGEM1 occupancy sites genome-wide (Supplementary Fig. 3d, e) and motif analysis uncovered a very significantly enriched AR response element (ARE) DNA motif (Supplementary Fig. 3f), revealing that ~82% of PCGEM1 co-localized with AR-bound sites (± 3 kb relative to the centre of PCGEM1 peak), of which ~70% corresponded to AR bound, H3K4me1-marked enhancers (Fig. 1f, g and Supplementary Fig. 3g), independently confirmed by quantitative (qPCR) analyses (Supplementary Fig. 3h, i) and ChIRP-Seq using even-numbered and odd-numbered probe sets (Supplementary Fig. 3j and data not shown). These data indicate a stoichiometry of PCGEM1 sufficient to account for the recruitment to AR DNA regulatory binding sites on enhancers. Levels of PRNCR1 in LNCaP cells are estimated as ~400–600 copies per cell (data not shown). The ability of these lncRNAs to read enhancer-associated histone marks might account for their preferential presence at AR-bound enhancers (see below).

By imposing a high stringency wash condition, we identified that DOT1L, CARM1, GADD45 α and AR specifically associated with *in vitro* transcribed biotinylated PRNCR1 by mass spectrometry analysis, whereas AR, β -catenin and pygopus 2 (PYGO2) preferentially associated with *in vitro* transcribed biotinylated PCGEM1 (Supplementary Fig. 4a–c; Supplementary Tables 1–3). β -catenin, CARM1 and GADD45 α have been suggested to have important roles in AR signalling⁷. Knockdown of AR, DOT1L, CTNNB1 and PYGO2 by specific siRNAs impaired DHT-induced activation of AR target genes, TMPRSS2, PSA (also known as KLK3) and FKBP5 (Supplementary Fig. 4d). Mass spectrometry analysis revealed that the lncRNA-bound AR contains several post-translational modifications, including K631/634 acetylation and K349 methylation (Supplementary Fig. 4e; Supplementary Tables 1–3). Consistent with the proposed importance of acetylation of AR in activation of an AR target gene⁸, a K631/634R mutation on AR inhibited its

¹Howard Hughes Medical Institute, Department of Medicine, University of California San Diego, La Jolla, California 92093, USA. ²Department of Molecular and Cellular Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. ³Department of Urology, School of Medicine, University of California Davis, Sacramento, California 95817, USA. ⁴Graduate Program, Kellogg School of Science and Technology, The Scripps Research Institute, La Jolla, California 92037, USA. ⁵Bioinformatics and System Biology Program, Department of Bioengineering, University of California San Diego, La Jolla, California 92093, USA. ⁶Neurosciences Graduate Program, Department of Biological Sciences, University of California San Diego, La Jolla, California 92093, USA.

*These authors contributed equally to this work.

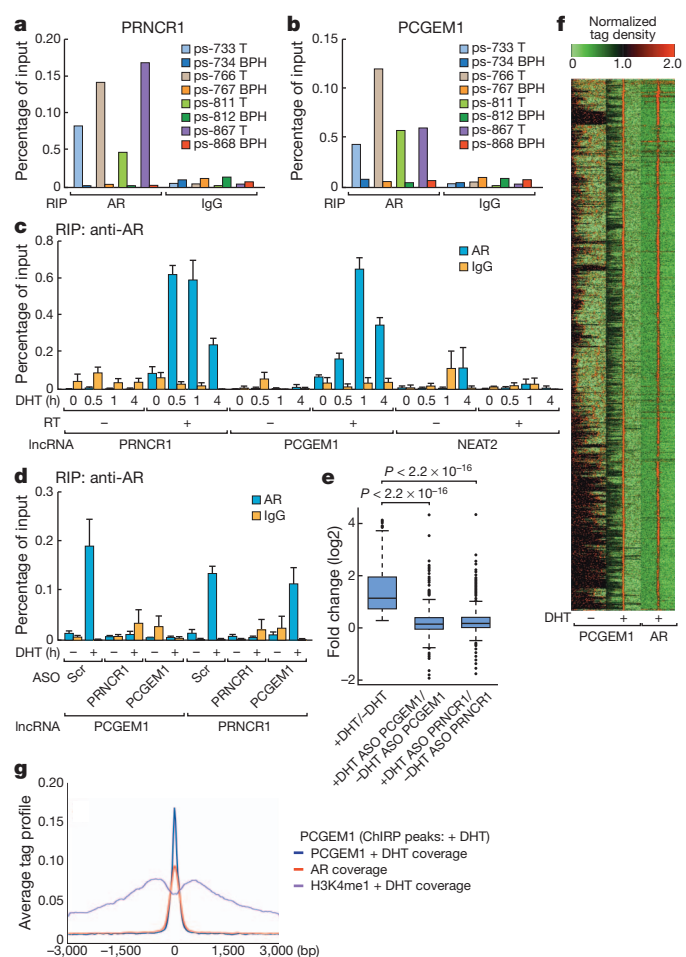


Figure 1 | Signal-dependent interaction between AR and prostate-specific lncRNAs. **a, b**, RIP assay performed in paired benign prostatic hyperplasia (BPH) and prostate tumour (T) tissues. **c**, RIP assay in DHT-treated LNCaP cells (100 nM) at the indicated time points. **d**, RIP assay in LNCaP cells transfected with indicated ASO followed by DHT treatment (100 nM). Scr, scrambled control. **e**, Global changes in DHT-induced AR target genes in PCGEM1 or PRNCR1 depleted LNCaP cells. **f**, Heatmap showing the distribution of PCGEM1 and AR binding sites in DHT-stimulated LNCaP cells. **g**, Average tag profile analysis of the aligned 2,142 PCGEM1 ChIP-seq peaks. Mean \pm s.e.m. for **c** and **d** ($n = 3$).

interaction with PRNCR1 and PCGEM1 (Fig. 2a; Supplementary Fig. 5a) and DHT-induced expression of AR target genes (Supplementary Fig. 5b), whereas overexpression of an AR(K631/634Q) mutant resulted in enhanced DHT-dependent interactions with PRNCR1 and PCGEM1 (Fig. 2a, b; Supplementary Fig. 5c). These data indicate that PRNCR1 and PCGEM1 interact with AR in a K631/634 acetylation- and K349 methylation-dependent manner, respectively.

Because effective AR–PCGEM1 interaction requires the methylation of AR at K349 (Fig. 2b), we confirmed DOT1L-mediated AR methylation at K349 using *in vitro* methylation assays, finding that a K349R point mutation markedly inhibited AR methylation (Supplementary Fig. 6a; Fig. 2c). Specific *DOT1L* knockdown impaired the interaction between AR and PCGEM1, but not that with PRNCR1 (Supplementary Fig. 6b, c), indicating that AR K349 methylation, mediated by PRNCR1-bound DOT1L, is critical for the recruitment of PCGEM1 to AR. Indeed, overexpression of an AR(K349R) mutant significantly reduced DHT-induced gene activation in LNCaP cells (Supplementary Fig. 5b).

In vitro binding studies demonstrated that PRNCR1 bound to the region of amino acids 549–623 of AR and PCGEM1 bound to the AR N-terminal region when methylated at K349 by overexpressed DOT1L

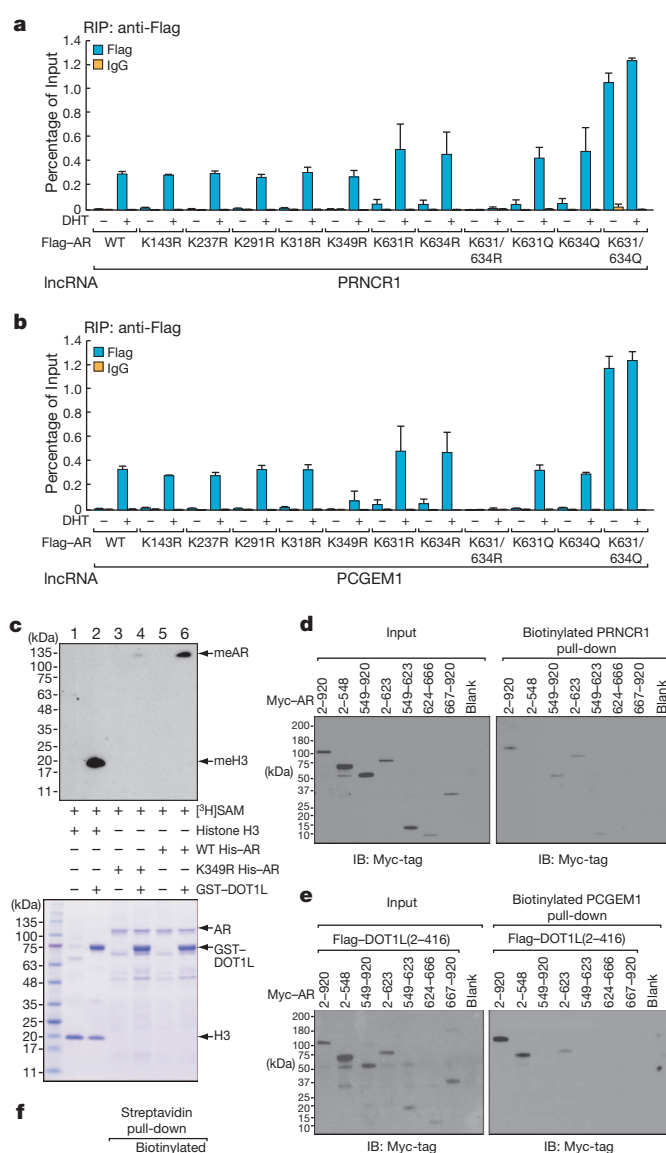


Figure 2 | Mechanistic study of lncRNA with associated transcription factors/co-activators. **a, b**, RIP assay in DHT-treated LNCaP cells (100 nM) expressing the indicated plasmids. **c**, *In vitro* methylation assay for AR. **d, e**, LNCaP cells expressing Myc-tagged AR fragments (**d**) or co-transfected with Flag-tagged DOT1L (**e**) were subjected to RNA pull-down assay. **f**, *In vitro* transcribed PCGEM1 full-length, (Δ 411–490), or (Δ 1191–1270) were incubated with cell lysates extracted from DHT-treated LNCaP cells (100 nM, 1 h) for an *in vitro* RNA pull-down assay. Mean \pm s.e.m. for **a** and **b** ($n = 3$).

(Fig. 2d, e; Supplementary Fig. 7a). By incubating *in vitro* transcribed PCGEM1 with nuclear lysate from cells overexpressing Myc-tagged PYGO2 proteins, including full-length, N- or C-terminally truncated proteins, we identified strong interactions between PCGEM1 and the PYGO2 C terminus (Supplementary Fig. 7b).

To map the sequence motif of PCGEM1 responsible for AR or PYGO2 binding, we performed modified *in vitro* RNA pull-down followed by dot-blot assay (Supplementary Fig. 8a), using two regions of NEAT2 bound by unmethylated PC2 (also known as CBX4) as a

control for the crosslinking and immunoprecipitation (CLIP) assay⁹ (Supplementary Fig. 8b). Methylated AR bound/protected PCGEM1 sequence was identified to encompass 421-GAT...TCC-480 (Supplementary Fig. 8c), with unmethylated AR or the unrelated protein His-tagged MURF1 (also known as TRIM63) not showing specific binding to any region of PCGEM1 (Supplementary Fig. 8c). A sequence motif of PCGEM1 encompassing 1201-TGT...ATT-1260, distinct from the AR binding region, was identified as the PYGO2 binding motif, with deletion of this motif (Δ 1191–1270) abolishing PYGO2 binding with no effect on AR binding (Supplementary Fig. 8c and Fig. 2f). Similarly, deletion of the AR binding site of PCGEM1 (Δ 411–490) abolished the AR–PCGEM1 interaction, with minimal effect on the PYGO2–PCGEM1 interaction (Fig. 2f).

Modified histone peptide array experiments using *in vitro* transcribed biotinylated PCGEM1 or PRNCR1 revealed that they selectively recognize H3K4me1 and H4K16ac histone marks indicative of enhancers^{10,11}, respectively (Supplementary Fig. 9a–c). Therefore, it is probable that these histone tail associations of PRNCR1 and PCGEM1 serve as a functional component of their preferred recruitment to enhancers of AR-regulated transcription units.

PCGEM1 and PRNCR1 were highly upregulated in the LNCaP-cds2 and CWR22Rv1 castration-resistant prostate cancer cell line models compared with immortalized ‘normal’ prostate epithelial cell lines, WPE and RWPE, or even LNCaP cells (Supplementary Fig. 10a, b). The AR antagonist, bicalutamide (Casodex), reduced the DHT-induced PSA expression in LNCaP cells but failed to act as an antagonist in LNCaP-cds2 cells (Supplementary Fig. 10c). Transduction of LNCaP-cds2 cells with lentivirus expressing short hairpin RNAs (shRNAs) against PRNCR1 or PCGEM1, but not a nonspecific (*lacZ*) shRNA, significantly reduced the activation of several canonical AR target genes but had no effect on AR expression levels (Supplementary Figs 10d–f and 11a). Because truncated forms of AR that exhibit ligand-independent transcriptional activation are frequently detected in castration-resistant prostate cancer and may often reflect alterations in AR gene structure, we investigated the potential roles of PRNCR1 and PCGEM1 in AR-mediated basal transcription activity in castration-resistant prostate cancer cells. RT-PCR (PCR with reverse transcription) using primers specific for one AR ‘splicing’ variant, which encodes a truncated length AR known as AR-V7 (ref. 12) confirmed the presence of this variant in LNCaP-cds2 cells (Supplementary Fig. 11b). Western blot analysis using N-terminal AR-specific antibody (441), revealed that the AR-V7 variant (~75 kilodaltons (kDa)) represents ~1–2% of total AR in LNCaP-cds2 cells, although it is the predominant form in CWR22Rv1 cells (Supplementary Fig. 10g). Because overexpression of truncated AR can constitutively activate androgen-responsive genes in the absence of ligand¹², we transfected LNCaP cells with the AR(Q641X) mutant, with activation of canonical androgen-responsive genes, including *TPMRSS2*, *PSA*, *KLK2*, *FKBP5* and *NKX3-1* in the absence of added androgen (Fig. 3a; Supplementary Fig. 12a, b). This constitutive effect of AR(Q641X) was highly reduced upon knockdown of either PRNCR1 or PCGEM1 (Fig. 3a). An RNA immunoprecipitation (RIP) assay in CWR22Rv1 cells demonstrated that both PRNCR1 and PCGEM1 associated constitutively with truncated AR (AR-V7) with or without ligand (Fig. 3b). By immunoblotting the AR-V7 immunoprecipitates with N-terminal AR-specific antibody (441), we did not observe any residual pull-down indicative of interaction between full-length and truncated AR (Fig. 3b, right panel), arguing against an indirect association of PRNCR1 and PCGEM1 with truncated AR because of heterodimerization with full-length AR. Using an antibody specific for the C-terminal ligand binding domain of AR (C-19) to selectively recognize full-length AR, we observed interactions between these lncRNAs and full-length AR in the absence of added ligand (Fig. 3b), possibly due to the relative higher level of basal acetylation and methylation of full-length AR in CWR22Rv1 cells (Supplementary Fig. 12c).

To study the biological roles of PRNCR1 or PCGEM1, we generated stable cell lines derived from CWR22Rv1 harbouring doxycycline

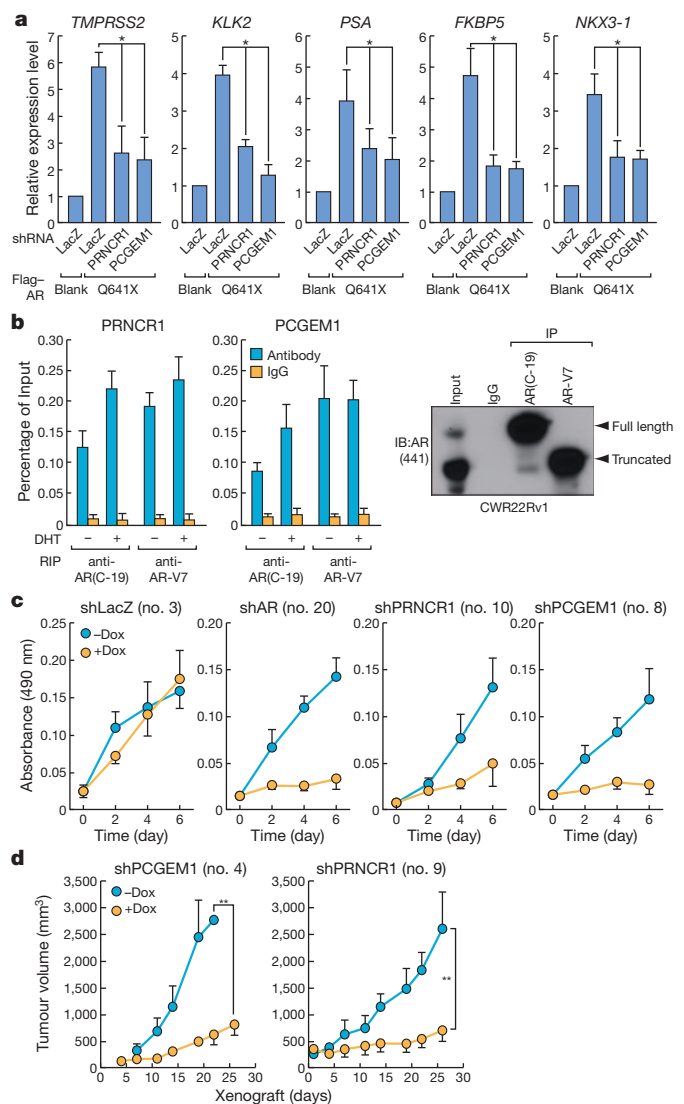


Figure 3 | PCGEM1 and PRNCR1 promote hormone-independent activation of the AR transcriptional program in castration-resistant prostate cancer. **a**, qRT-PCR analyses of AR targets in LNCaP cells co-transfected with indicated vectors followed by doxycycline induction (160 ng ml⁻¹, 2 days). **b**, RIP assay in CWR22Rv1 cells treated with or without DHT (100 nM, 1 h) using the indicated antibodies. Detection of immunoprecipitated full-length AR and AR-V7 were shown. **c**, Cell proliferation assay in CWR22Rv1 cells stably expressing indicated shRNAs followed by doxycycline induction (160 ng ml⁻¹) for the indicated times. $P \leq 0.01$ for the last three panels. **d**, Xenografts of CWR22Rv1 cell lines harbouring doxycycline-induced shRNAs were monitored for tumour growth for the indicated time, with or without doxycycline intake (4 mice per group). Mean \pm s.e.m. ($n = 6$, * $P < 0.05$ and ** $P < 0.01$).

(dox)-induced shRNA against *lacZ*, PRNCR1 or PCGEM1 (Supplementary Fig. 13a). Dox-induced either PCGEM1 or PRNCR1 knockdown significantly reduced the expression of canonical AR target genes, with no noticeable effect on AR expression level (Supplementary Figs 11a, 13b). Dox-induced knockdown of either PRNCR1 or PCGEM1 also inhibited the growth of CWR22Rv1 cells comparable to the effect of AR knockdown¹³, without affecting AR expression levels (Fig. 3c; Supplementary Figs 11a and 13c). Remarkably, conditional shRNA-mediated inhibition of either PRNCR1 or PCGEM1 robustly inhibited *in vivo* tumour growth in a CWR22Rv1 prostate cancer xenograft mouse model (Fig. 3d), indicative of an lncRNA-dependent regulatory network that critically regulates growth of castration-resistant prostate cancer cells *in vivo*.

Although knockdown of either PRNCR1 or PCGEM1 did not affect the recruitment of AR on PSA and KLK2 enhancers (Supplementary

Fig. 14a, b, left panels), knockdown of *PCGEM1* inhibited SMC1 recruitment on *PSA* and *KLK2* promoters, with only minimal effects on SMC1 levels on enhancers (Supplementary Fig. 14a, b, right panels), consistent with proposed cohesin-dependent¹⁴ formation of chromatin loops between enhancers and promoters. We further demonstrated ligand-induced enhancer–promoter interactions in the *PSA* transcription unit by chromatin immunoprecipitation (ChIP)–3C (3C, chromosome conformation capture) assay¹⁵ and found that these interactions were impaired by depletion of either *PRNCR1* or *PCGEM1* (Fig. 4a; Supplementary Fig. 14c).

The ability of PYGO2, associated with PCGEM1 (Supplementary Fig. 2a), to recognize a canonical promoter histone mark¹⁶, H3K4me3, raised the possibility that PYGO2 might be involved, at least quantitatively, in enhancer–promoter looping. ChIP assays revealed that PYGO2 was efficiently recruited to enhancer and promoter regions of the *PSA*, *KLK2* and *TMPRSS2* transcription units in a DHT-dependent manner, exhibiting relatively higher and earlier association with the enhancer regions (Fig. 4b; Supplementary Fig. 15a). Knockdown of either *PRNCR1* or *PCGEM1* in LNCaP cells inhibited PYGO2 recruitment to AR-dependent

enhancers/promoters (Fig. 4c; Supplementary Fig. 15b). On knockdown of *PYGO2*, AR and SMC1 failed to effectively associate with the *PSA*, *KLK2* or *TMPRSS2* promoters despite unperturbed DHT-dependent AR or SMC1 recruitment to their enhancers (Supplementary Fig. 16a–d) and DHT-induced enhancer–promoter looping in the *PSA* transcription unit was inhibited (Fig. 4d). GRO-Seq analysis revealed a broad inhibition of AR-dependent transcriptional program under the condition of *PYGO2* knockdown ($n = 290$, edgeR FDR < 0.01), which did not affect *AR* expression (Fig. 4e; Supplementary Fig. 16a).

Depleting *PRNCR1* or *PCGEM1* from LNCaP-cds2 cells also inhibited enhancer–promoter looping in *FASN* and *NDRG1* (Supplementary Fig. 17a–d; Fig. 4f), previously shown to be activated following FoxA1 knockdown in LNCaP cells¹⁷. Again, knockdown of either *PRNCR1* or *PCGEM1* had no effect on recruitment of AR to enhancer regions of the *FASN* and *NDRG1* transcription units or on *PYGO2* expression levels (Supplementary Fig. 17e).

To address whether the PYGO2 PHD domain might itself be instrumental for its function in mediating chromatin looping, we first depleted PYGO2 by shRNA treatment followed by overexpression in

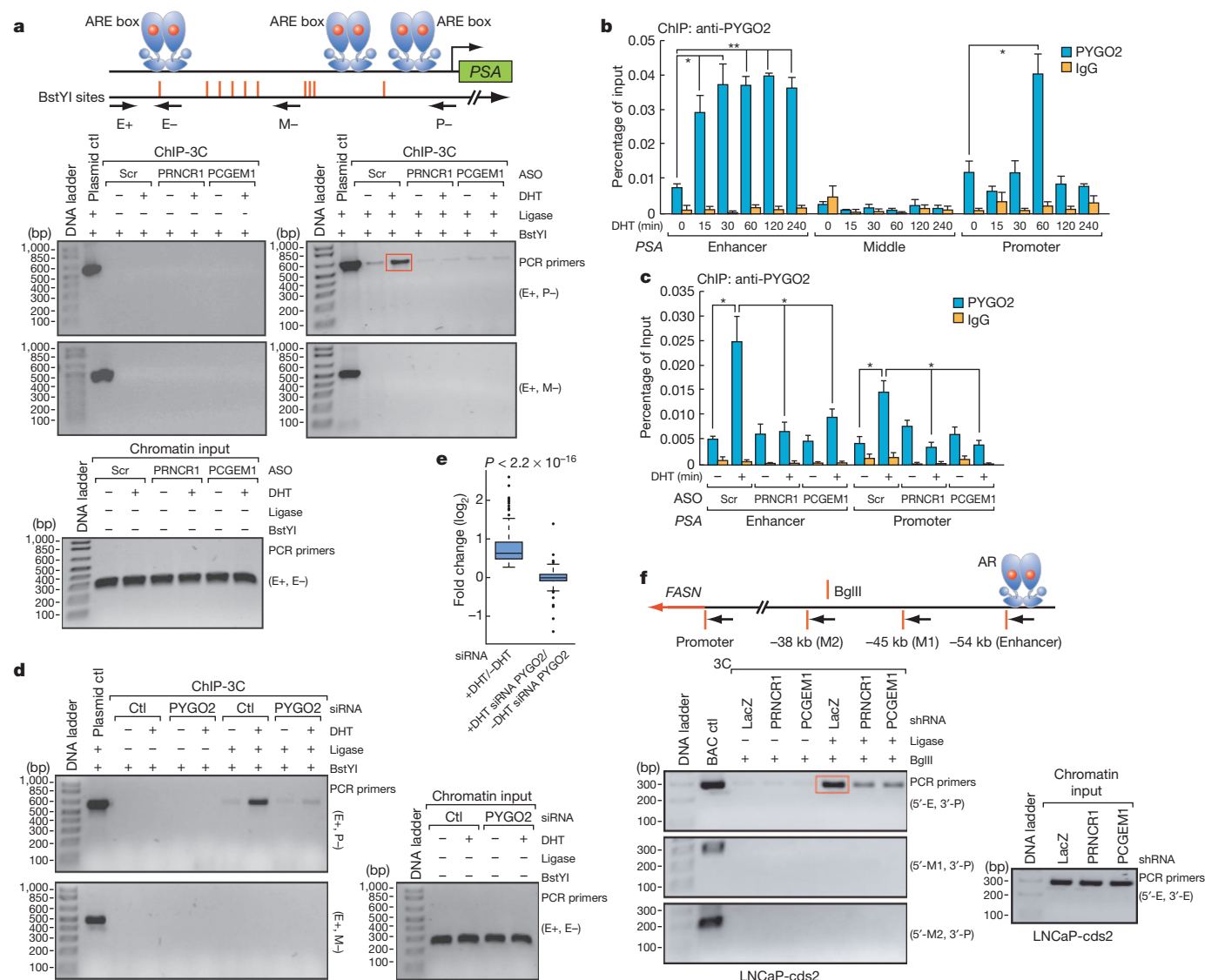


Figure 4 | Regulation of enhancer–promoter interaction by *PRNCR1* and *PCGEM1*. **a**, ChIP–3C in LNCaP cells with indicated ASO transfection and treatment. Ctl, control. **b**, ChIP–qPCR showing PYGO2 occupancy in DHT-treated LNCaP cells (100 nM) at indicated times. **c**, ChIP–qPCR indicate the PYGO2 recruitment in LNCaP cells with indicated transfection and treatment. **d**, ChIP–3C in LNCaP cells with indicated siRNA transfection and treatment.

e, Global changes in DHT-induced AR targets in LNCaP cells with indicated siRNA transfection and treatment. **f**, Chromosome conformation capture (3C) assay on the *FASN* locus in LNCaP-cds2 cells transduced with indicated shRNA harbouring lentivirus. The red box indicates the PCR product sequenced in **a** and **f**. Mean \pm s.e.m. ($n = 3$, $*P < 0.05$ and $**P < 0.01$).

LNCAp-cds2 cells of either shRNA-resistant wild-type PYGO2 (ref. 18) or a W352A mutant defective for H3K4me3 recognition¹⁶. In 3C assays, knockdown of *PYGO2* reduced *FASN* enhancer–promoter interactions, which could be rescued by overexpression of wild-type, but not W352A, PYGO2 (Supplementary Figs 18a and 19a), even though there was equal recruitment of wild-type PYGO2 or the W352A mutant to enhancers, and no altered promoter H3K4me3 levels (Supplementary Fig. 18b–d). Knockdown of *PYGO2* curtailed the expression of canonical AR target genes *TMPRSS2*, *KLK2*, *PSA*, *FKBP5* and *NKX3-1*, and overexpression of wild-type, but not W352A, PYGO2 was able to robustly rescue the induction of these genes (Supplementary Fig. 19b). These data indicate that PYGO2 exerts a quantitatively important role in DHT-dependent enhancer–promoter interactions and coding target gene activation. For 220 AR-regulated coding gene promoters under regulation of an enhancer we observed recruitment of PYGO2 by ChIP followed by high-throughput sequencing (ChIP-Seq), but did not observe ligand-induced increase in the next adjacent, non-AR-regulated transcription unit (~204 promoters) (Supplementary Fig. 19c); indicating that the H3K4me3 mark cannot alone be sufficient to effectively recruit PYGO2 and suggesting a role for other similarly modified proteins in prostate cancer cell gene activation events.

In the present study, we have found a mechanistic link between prostate cancer-upregulated lncRNAs and AR transcriptional activity, revealing the biological importance of the lncRNAs, *PRNCR1* and *PCGEM1*, in licensing C-terminally truncated, as well as full-length, AR-dependent gene activation events in prostate cancer cells (Supplementary Fig. 19d). Considering the regulatory potential of enhancer-derived RNAs identified in recent studies^{19,20}, lncRNAs may also be part of a broad transcription regulatory network.

METHODS SUMMARY

Experiments were performed in LNCAp, LNCAp-cds1, LNCAp-cds2, CWR22Rv1, WPE and RWPE cells. ChIRP-Seq, GRO-Seq, ChIP-Seq, ChIP-3C and 3C assays were performed as described^{6,15,17} with modifications (details in Methods section). All siRNAs used in this study were either validated by vendors (Dharmacon, Qiagen and Sigma-Aldrich) or by our in-house RT–qPCR/western blotting analyses for specificity (Supplementary Figs 6b and 16a) and used only if providing >70% knockdown efficiency. Antisense oligonucleotides were designed and synthesized by IDT, with knockdown specificity and efficiency of 3 individual sequences tested for *PRNCR1* or *PCGEM1* (Supplementary Fig. 2a). shRNAs against *PRNCR1* and *PCGEM1* were designed at <http://biosettia.com/support/shrna-designer>, with knockdown specificity and efficiency of 2 individual sequences tested for *PRNCR1* or *PCGEM1* (Supplementary Fig. 10d). ChIRP probes were designed at <http://www.singlemoleculefish.com/designer.html>. Relative quantities of gene expression level were normalized to the actin gene. The relative quantities of ChIP samples were normalized by individual inputs, respectively. Results are reported as mean \pm s.e.m. of three or six independent experiments as indicated. Comparisons were performed using a two-tailed paired Student's *t*-test.

Full Methods and any associated references are available in the online version of the paper.

Received 4 December 2012; accepted 12 June 2013.

Published online 14 August 2013.

1. Gupta, R. A. *et al.* Long non-coding RNA *HOTAIR* reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071–1076 (2010).
2. Heinlein, C. A. & Chang, C. Androgen receptor in prostate cancer. *Endocr. Rev.* **25**, 276–308 (2004).

3. Scher, H. I. & Sawyers, C. L. Biology of progressive, castration-resistant prostate cancer: directed therapies targeting the androgen-receptor signaling axis. *J. Clin. Oncol.* **23**, 8253–8261 (2005).
4. Petrovics, G. *et al.* Elevated expression of *PCGEM1*, a prostate-specific gene with cell growth-promoting function, is associated with high-risk prostate cancer patients. *Oncogene* **23**, 605–611 (2004).
5. Chung, S. *et al.* Association of a novel long non-coding RNA in 8q24 with prostate cancer susceptibility. *Cancer Sci.* **102**, 245–252 (2011).
6. Chu, C., Qu, K., Zhong, F. L., Artandi, S. E. & Chang, H. Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA–chromatin interactions. *Mol. Cell* **44**, 667–678 (2011).
7. Kypta, R. M. & Waxman, J. Wnt/ β -catenin signalling in prostate cancer. *Nature Rev. Urology* **9**, 418–428 (2012).
8. Fu, M. *et al.* Acetylation of androgen receptor enhances coactivator binding and promotes prostate cancer cell growth. *Mol. Cell. Biol.* **23**, 8563–8575 (2003).
9. Yang, L. *et al.* ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. *Cell* **147**, 773–788 (2011).
10. Zippo, A. *et al.* Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation. *Cell* **138**, 1122–1136 (2009).
11. Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
12. Hu, R. *et al.* Ligand-independent androgen receptor variants derived from splicing of cryptic exons signify hormone-refractory prostate cancer. *Cancer Res.* **69**, 16–22 (2009).
13. Sun, A. *et al.* Adeno-associated virus-delivered short hairpin-structured RNA for androgen receptor gene silencing induces tumor eradication of prostate cancer xenografts in nude mice: a preclinical study. *Int. J. Cancer* **126**, 764–774 (2010).
14. Taberlay, P. C. *et al.* Polycomb-repressed genes have permissive enhancers that initiate reprogramming. *Cell* **147**, 1283–1294 (2011).
15. Wang, Q., Carroll, J. S. & Brown, M. Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. *Mol. Cell* **19**, 631–642 (2005).
16. Gu, B. *et al.* Pygo2 expands mammary progenitor cells by facilitating histone H3 K4 methylation. *J. Cell Biol.* **185**, 811–826 (2009).
17. Wang, D. *et al.* Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474**, 390–394 (2011).
18. Gu, B., Watanabe, K. & Dai, X. Pygo2 regulates histone gene expression and H3 K56 acetylation in human mammary epithelial cells. *Cell Cycle* **11**, 79–87 (2012).
19. Lam, M. T. Y. *et al.* Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature* **498**, 511–515 (2013).
20. Li, W. *et al.* Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* **498**, 516–520 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to X. Dai for providing the PYGO2 shRNA and cDNA constructs and J. Hightower for assistance with figure presentation. This work was supported by National Institutes of Health (NIH) grants DK039949, DK18477, NS034934 and CA173903, Department of Defense grant and initially by a grant from Prostate Cancer Foundation to M.G.R.; by Department of Defense grant PC111467 and SV2C-AACR-DT0812 to C.D.E.; by grants from the NIH Pathway to Independence Award (1K99DK094981–01) to C.-R.L.; by US Army Medical Research and Material Command Era of Hope Postdoctoral award (W81XWH-08-1-0554), NIH Pathway to Independence Award (4R00CA166527–02) and Cancer Prevention Research Institute of Texas First-time Faculty Recruitment Award (R1218) to L.-Q.Y.; C.-Y.J. is the recipient of a Cancer Research Institute Postdoctoral Fellowship. M.G.R. is an Investigator of the Howard Hughes Medical Institute.

Author Contributions L.-Q.Y., C.-R.L. and M.G.R. designed the research, and L.-Q.Y. and C.-R.L. performed most of the experiments, with participation from C.-Y.J.; J.C.Y., under supervision of C.P.E., performed *in vivo* tumour xenograft experiments. B.T. and D.Mer. performed bioinformatics analyses on GRO-Seq, ChIP-Seq and ChIRP-Seq data. W.-B.L., J.Z. and K.A.O. conducted high-throughput sequencing, and D.Men. helped with ChIRP assays. L.-Q.Y., C.-R.L. and M.G.R. wrote the manuscript.

Author Information The high-throughput sequencing data sets are deposited in the Gene Expression Omnibus database under accession GSE47807. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.G.R. (mrosenfeld@ucsd.edu), L.-Q.Y. (lyang7@mdanderson.org) or C.-R.L. (clin2@mdanderson.org).

METHODS

Tissue samples and processing. Experiments using paired benign prostatic hyperplasia (BPT) and tumour (T) were performed from unidentified individual prostate cancer patients, which were obtained from R. W. deVere White, UC Davis Comprehensive Cancer Center. Informed consent was obtained from all subjects. The tissue samples used in this manuscript were received as de-identified samples without any protected health information (PHI) attached. The Gleason score or tumour/BPH status was considered pathological information, not patient information. We did not know the sample donors' names or birth dates. The tissues were homogenized by Precellys24 tissue homogenizer followed by downstream assays.

Cell culture and transfection. Prostate cancer LNCaP cells were obtained from ATCC and cultured in RPMI1640 containing 10% (vol/vol) FBS. The benign immortalized prostate cell line RWPE, WPE and the castration-resistant prostate cancer cell lines LNCaP-cds1, LNCaP-cds2, CWR22Rv1 were provided by C. Evans. RWPE and WPE cells were grown in KGM media and BulletKit from Lonza supplemented with L-Glutamine. CWR22Rv1, LNCaP-cds1 and LNCaP-cds2 cells were grown in RPMI1640 media containing 5% final volume of charcoal stripped serum. LNCaP cells were grown to 30–50% confluence and siRNA/ASO transfections were carried out using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Transfection of LNCaP cells with DNA plasmids was performed using Amaxa Nucleofector Kit R from Lonza. shRNAs specific for *lacZ*, *PRNCR1*, *PCGEM1* or AR were delivered, by lentiviral transduction, to LNCaP-cds2 and CWR22Rv1 cells.

Cloning procedures. The full-length AR expression vector has been previously described¹⁷. Human *PCGEM1* (14–1556) and *PRNCR1* fragments (1–3240, 3156–6428, 6331–9670 and 9531–12710) were amplified from complementary DNAs (cDNAs) generated from LNCaP cells and cloned into pSTBlue-1 vector (Novagen) for *in vitro* transcription assay. The *PCGEM1* gene sequence with 411–490 and 1191–1270 deletion were synthesized by GeneScript and cloned into pcDNA3.1 vector (Invitrogen). Lentiviral vector pLKO.1 containing the shRNA against scrambled sequence, PYGO2 and pHIV vector containing RNAi-resistant PYGO2 cDNA were obtained from X. Dai (Department of Biological Chemistry, University of California at Irvine)¹⁸. A 4.8 kb genomic sequence upstream of the *PSA* promoter was amplified from LNCaP cells and subcloned into pSTBlue-1 vector (Novagen). Bacterial expression vectors for AR and K349R mutant were constructed by subcloning the gene sequences into pET-28a backbone (Novagen). Nuclear expression vectors for AR(2–920), AR(2–548), AR(549–920), AR(2–623), AR(549–623), AR(624–666), AR(667–920) and PYGO2(2–406), PYGO2(2–266), PYGO2(267–406) were constructed by subcloning the cDNA sequences into pCMV/myc/nuc backbone (Invitrogen). The expression vector of Flag-DOT1L was provided by Y. Zhang's laboratory. All mutants were generated using QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent Technologies). Detailed oligonucleotide sequences are listed in the oligonucleotide sequences and primers section.

Antisense DNA oligonucleotide, siRNA and lentiviral shRNA. 2'-O-methyl phosphorothioate oligonucleotides were designed and synthesized by Integrated DNA Technologies. Commercially available FlexiTube siRNA targeting AR (SI02757258) and CARM1 (SI02663815) from Qiagen, ON-TARGETplus SMARTpool siRNA targeting DOT1L (L-014900–01–0005) and GADD45α (L-003893–00–0005) from Dharmacon, MISSION siRNA targeting β-catenin (SASI_Hs01_00117958), Pygopus 2 (SASI_Hs01_00059018) from Sigma-Aldrich were used in this study. The knockdown efficiency and specificity of all siRNAs were validated either by vendors or ourselves (Supplementary Fig. 6b, 16a). Oligonucleotides for shRNA targeting *PRNCR1*, *PCGEM1* and AR were designed at (<http://biosettia.com/support/shrna-designer>) and cloned into pLV-H1TetO-GFP-Puro vector according to the manufacturer's instructions (Biosettia). We tested the efficacy and specificity of 3 ASOs (see Supplementary Fig. 2a) and 2 shRNAs (see Supplementary Fig. 10d) in LNCaP and LNCaP-cds2 cells, respectively, for both *PRNCR1* and *PCGEM1*. For functional assays, the specific ASO/shRNA giving the best knockdown efficiency was used. Detailed ASO/shRNA sequences are listed in the oligonucleotide sequences and primers section.

Antibodies. Specific antibodies were purchased from the following commercial sources: anti-AR (N-20), anti-AR (C-19), anti-AR (441), anti-GR (E-20), anti-β-catenin (D-10) and anti-GAPDH (6C5) from Santa Cruz Biotechnology; anti-CARM1 (4438), anti-GADD45α (3518) and anti-pan acetylated-lysine (#9441) from Cell Signaling Technology; anti-Flag M2, anti-γ-Tubulin (T5326) and anti-PYGO2 from Sigma-Aldrich Prestige Antibodies; anti-PYGO2 (GTX119726) from GeneTex; Anti-DOT1L (39954) and anti-Myc tag (clone 4E12) from Active Motif; anti-SMC1 (A300–055A) from Bethyl Laboratories; anti-pan methylated lysine (7315) from Abcam and anti-AR-V7 from Precision Antibody.

Protein recombination and purification. Recombinant His-AR proteins were expressed in *E. coli* strain BL21-CodonPlus (DE3)-RIPL (Agilent Technologies) and purified using TALON Metal Affinity Resin (Clontech). Recombinant histone H3 was purchased from Active Motif. Human DOT1L (amino acids 2–416) was

obtained from Sigma-Aldrich. Human PYGO2 was purchased from BioClone. His-tagged MURF1 was purchased from BostonBiochem.

RIP. RIP was performed in native conditions as described²¹. Briefly, 1×10^7 LNCaP cell nuclei were pelleted and lysed in 1 ml ice-cold polysomal lysis buffer (100 mM KCl, 5 mM MgCl₂, 10 mM HEPES pH 7.0, 0.5% NP-40, 1 mM DTT (dithiothreitol)) supplemented with anti-RNase, protease inhibitor cocktail, phosphatase inhibitor cocktail, panobinostat and methylstat. The lysate were passed through a 27.5 gauge needle 4 times to promote nuclear lysis. Turbo DNase (400 U) was then added to the lysate and incubated on ice for 30 min. The cell lysate was diluted in the NT2 buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM MgCl₂, 0.05% NP-40) and 50 µl of the supernatant was saved as input for PCR analysis. 500 µl of the supernatant was incubated with 4 µg of AR (441) antibodies-IgG magnetic beads (pre-blocked by $1 \times$ PBS + 5 mg ml⁻¹ BSA) at 4 °C overnight. The RNA/antibody complex was washed four times (1 ml wash, 5 min each) by NT2 buffer supplemented with anti-RNase, protease inhibitor cocktail, phosphatase inhibitor cocktail, panobinostat and methylstat. The RNA was extracted using acid-phenol: chloroform, pH 4.5 (with IAA, 125:24:1) (Invitrogen) according to the manufacturer's protocol and subjected to RT-qPCR analysis.

ChIRP. ChIRP was performed as described⁶ with minor modifications. Briefly, 60-mer antisense DNA probes targeting *PRNCR1* and *PCGEM1* full-length sequences were designed at <http://www.singlemoleculefish.com/designer.html>. A set of probes targeting *lacZ* RNA was also designed as the negative control. All probes were biotinylated using Label IT nucleic acid biotin labelling kit from Mirus Biotechnology. LNCaP cells were fixed with 1% formaldehyde for 10 min at room temperature. Crosslinking was then quenched with 125 mM glycine for 5 min. The chromatin preparation, hybridization/elution, deep sequencing steps were essentially performed as described⁴ except that the wash was conducted at 50 °C and 65 °C. The image analysis and base calling were performed using Illumina's computational analysis pipeline. The sequencing reads were aligned to hg18 human genome by using Bowtie2 (ref. 22) and only one read per position was kept for downstream analyses. Peak finding was performed by using HOMER suite²³ and the peaks within 1 kb from each other were merged. Peak intersection was computed by using intersectBed in BedTools²⁴, after extending the peaks with 1 kb. In order to call reliable peaks, we have excluded from analysis the peaks that overlapped the satellite repeats or *lacZ* ChIRP peaks. The annotation of the peaks on the human genome and the tag density profiles were computed in HOMER, and the display of the heatmap were carried in MeV²⁵. Sequenced motif enrichment analysis was carried in HOMER. For ChIP-Seq data (AR and H3K4me1), peak finding was performed by using HOMER or SICER²⁶.

Data processing procedure for overlapping ChIP-Seq and ChIRP-Seq. We used the standard ChIP-Seq peak finding tools in the processing of ChIRP-Seq data. We chose HOMER software suite (<http://biowhat.ucsd.edu/homer/>) for the analysis of both AR ChIP-Seq and *PCGEM1* ChIRP-Seq data, using the same program routine (that is, findPeaks command) and the same default parameters that calls only the robust peaks (these parameters are outlined below): (1) fold enrichment over input tag count, default: 4.0; (2) Poisson *P* value threshold relative to input tag count, default: 0.0001; (3) fold enrichment over local tag count, default: 4.0; (4) *fdr* <#> false discovery rate, default = 0.001; and (5) size of region used for local filtering = 10000. Peak finding procedure: (1) typically, the tag distribution along the genome could be modelled by a Poisson distribution and the peak finding algorithm slides windows of fixed size across the genome to find candidate peaks with a significant tag enrichment (Poisson distribution *P* value default 10–4 to 10–5); (2) we use very strict parameters in HOMER to call a peak: a very low FDR (that is, 0.001), and a high fold enrichment over input tag count (that is, 4). By default, HOMER also requires the tag density at peaks to be fourfold greater than in the surrounding 10 kb region. Therefore, we ensure that only the sharp peaks with low local background are called and considered for downstream analyses. Both ChIP-Seq and ChIRP-Seq data were processed precisely in the same way using the same default settings of HOMER.

For the analysis of overlapping between ChIP-Seq and ChIRP-Seq data, we used the following samples: *PCGEM1* ChIRP-Seq (–DHT), *PCGEM1* ChIRP-Seq (+DHT) and AR ChIP-Seq (at a higher sequencing depth). The heatmaps were generated in two steps. First, we used HOMER and the command (annotatePeaks.pl < peak file > < genome > -size <#> -hist <#> -ghist) that specifies the list of peaks (< peak file >), and the tag density in a region that covers < size > kb around the peak centre; the tag density is specified in a bin size that is specified by the parameter < hist > to generate a matrix of tag densities across the samples, for each peak. Typically the tag densities are normalized to 10 million reads for each sample. Second, we used MeV package (<http://www.tm4.org/mev.html>) to display the heatmap and to colour code it on a scale from 0 to 2: a difference in the colours from 2 (red) to 0 (green) may reflect a fold enrichment over local tag count.

GRO-Seq. The image analysis and base calling were performed by using Illumina's standard computational analysis pipeline. Bowtie2 (ref. 22) was used to align the

sequencing reads to hg18 human genome; when multiple reads aligned to the same genomic position, only one read per position was kept for downstream analyses. We have used RefSeq annotations in order to estimate the gene expression level by counting the sequencing reads over the gene body (from 400 bp downstream of TSS to TTS) on the sense strand with respect to the gene transcription, by using intersectBed in BedTools²⁴. In order to call statistically significant differentially expressed genes, we have used edgeR²⁷ and a FDR <0.01. Additional criteria, including read density (normalized number of reads per kb) were imposed in order to filter the differentially expressed genes. Wilcoxon tests for sample comparisons were computed in R.

ChIP, ChIP-Seq and ChIP-3C. ChIP assays and ChIP-Seq was performed as described²⁸. ChIP-3C assays were performed as described¹⁵ with minor modifications. Briefly, LNCaP cells were cross-linked with 1% formaldehyde for 10 min followed by incubation with 125 mM glycine for 5 min. The cross-linked chromatin was sonicated and digested with the restriction enzyme BstYI (New England Biolabs) followed by immunoprecipitation with specific antibody coupled to Protein A Dynabeads (Invitrogen). The bound chromatin on the beads was ligated with T4 DNA ligase (Promega), eluted and reverse-crosslinked. The ChIP-3C material was purified by QIAquick gel extraction Kit (Qiagen) and subjected to PCR analysis with sets of primers as described¹⁵.

3C assay. LNCaP-cds2 cells were cross-linked with 1% formaldehyde for 10 min at room temperature and processed according to the procedures as described¹⁷ with minor modification. Briefly, BAC clones (Empire Genomics) for *FASN* (RP11-10331, RP11-10339) and *NDRG1* (RP11-671M3, RP11-125119) loci were used to generate control templates for the positive controls. 15 µg of the BAC clone for each locus were mixed and digested with 300 units of the corresponding restriction enzyme overnight at 37 °C. DNA fragments were extracted and ligated with T4 DNA ligase at a DNA concentration of 300 ng µl⁻¹. The primers for the fragments on the *FASN* and *NDRG1* loci were used as previously described¹⁷.

In vitro RNA pull-down coupled with dot-blot assay. *In vitro* transcribed biotinylated RNA was formed secondary structure as described²⁹ and incubated with recombinant His-tagged or GST-tagged proteins in gel shift protein binding buffer (Promega) on ice for 1 h. The reactions were ultraviolet (UV) irradiated (150 mJ per cm²) to crosslink RNA-protein complexes. After UV irradiation, the RNA was partially digested by RNaseI (Ambion) at 1:50 and 1:500 dilutions for 5 min, allowing a small fragment to remain attached to protein. RNA-protein complexes of interest were then partially purified by His tag or GST tag magnetic beads and the purified RNA-protein complexes were treated with proteinase K, which removes protein but leaves intact RNAs. The recovered RNAs were hybridized to BrightStar-plus positively charged nylon membrane spotted with 60-mer antisense DNA oligonucleotides tiling along the indicated lncRNA sequence at 37 °C overnight. The anti-sense oligonucleotides corresponding to the indicated lncRNAs were spotted on membrane as following order (left to right in each row): A1 corresponding to the oligonucleotide sequence nucleotides 1–60 of the indicated lncRNA, A2 corresponding to oligonucleotide sequence nucleotides 61–120 of indicated lncRNA and so on till the end of the lncRNA sequences. The hybridized membrane was washed as described at sequential 37 °C, 50 °C and 65 °C. The protein-bound RNA sequence was visualized by detection of Streptavidin-HRP signals. The positions and sequences of antisense DNA oligonucleotides tiling along *PCGEM1* is available upon request.

In vitro methylation assay. *In vitro* methylation assays were conducted in a total volume of 30 µl using 2 µg of substrate, and 1 µg of recombinant GST-DOT1L in methylation buffer (50 mM HEPES pH 8.0, 0.01% (v/v) NP-40, 10 mM NaCl, 1 mM DTT and 1 mM PMSF) supplemented with either 40 µM S-adenosyl-methionine (SAM, Sigma-Aldrich; for nonradioactive methylation) or 300 nM S-adenosyl-L-[methyl-³H]methionine ([³H]SAM, Perkin Elmer; 0.55 µCi per reaction for radioactive methylation). Reactions were carried out at 37 °C for 3 h and the reactions were separated on a 4–12% SDS-PAGE gel. The resulting protein bands were visualized by Coomassie blue staining, immunoblotting or autoradiography using ENHANCE spray (PerkinElmer). Purification of methylated protein was carried out as described⁹.

Cell lysis, protein immunoprecipitation and immunoblotting. Cells were homogenized in 1× RIPA buffer with protease inhibitor and anti-RNase (Invitrogen). Lysates were cleared by centrifugation at 17,000g for 15 min at 4 °C. The protein concentration of the extracts was determined by Bradford assay (Bio-Rad). Immunoprecipitation experiments and immunoblotting analyses were performed as described in previous studies⁹. The blotting signals were detected using SuperSignal West Dura extended duration substrate (Pierce).

Lentivirus packaging and transduction. Production of lentiviral shRNA stocks were performed in 293LTV cell line according to the manufacturer's instruction (Cell Biolabs). The lentivirus was further purified and concentrated by ViraBind Lentivirus Concentration and Purification Kit (Cell Biolabs). LNCaP-cds2 and CWR22Rv1 cells were transduced using ViraDuctin Lentivirus Transduction Kit (Cell Biolabs) and purified with 0.6 and 0.4 µg ml⁻¹ puromycin, respectively.

To establish stable cell lines expressing shRNA against *lacZ*, *PRNCR1*, *PCGEM1* or *AR*, CWR22Rv1 cells were transduced as described above and stable transduced clones were generated following selection with 0.4 µg ml⁻¹ puromycin. Single cell clones were then isolated by three rounds of single cell dilution, and the efficiency of doxycycline-induced target gene knockdown was screened by RT-qPCR with 24 individual clones for each target gene (see Supplementary Fig. 13a).

RNA pull-down assay and mass spectrometric analysis. RNA pull-down assay was performed as previously described⁹ with minor modifications. Biotin-labelled *PCGEM1* RNA (14–1556) and *PRNCR1* RNA fragments (1–3240, 3156–6428, 6331–9670 and 9531–12710) were *in vitro* transcribed with the Biotin RNA labelling mix (Roche) and T7 or SP6 RNA polymerase (Promega) and purified with RNeasy Mini Kit (Qiagen). RNAs were incubated with nuclear extract from LNCaP cells in the presence of anti-RNase, protease inhibitor cocktail, phosphatase inhibitor cocktail, panobinostat and methylstat. Proteins pulled down by biotinylated *PRNCR1* and *PCGEM1* were first eluted, denatured, reduced, alkylated and then digested with immobilized trypsin (Promega). The digested peptides were batch purified on a reversed-phase ZipTip (Millipore), and the resulting peptide pools were then subjected to mass spectrometric analysis at UCSD Biomolecular and Proteomics Mass Spectrometry Facility. Data were analysed using ProteinPilot software (Applied Biosystems).

Modified histone peptide array. The interaction between *in vitro* transcribed *PRNCR1* and *PCGEM1* with histone was analysed by Modified Histone Peptide Array (Active Motif) according to the manufacturer's instructions. The specificity of interactions was quantified by Array Analyses software (Active Motif).

Histone peptide pull-down assay. Histone peptide pull-down assays were performed using Sensolyte methylated histone H3 Kit from AnaSpec. Briefly, 1 µg of recombinant PYGO2 protein was incubated with 1 µg of biotinylated histone H3 peptides (unmodified, K4me1, K4me2, or K4me3) in the presence of yeast tRNA or *in vitro* transcribed *PCGEM1* in binding buffer (50 mM Tris-HCl, pH 7.5, 15 mM NaCl, 0.05% NP-40) for 2 h at 4 °C with rotation. After 1 h incubation with streptavidin magnetic beads and extensive washing, the bound protein was analysed by SDS-PAGE and western blotting with anti-PYGO2 antibodies. For this assay, *PCGEM1* was *in vitro* transcribed using unlabelled UTP.

Cell proliferation assay. Cell proliferation assay was performed using CellTiter 96 AQueous One Solution Cell Proliferation Assay (MTS) (Promega). Briefly, cells were distributed in 96-well plate with 100 µl media each. After removing the media and rinsed once with PBS, cells were supplied with 100 µl PBS mixed with 4 µl MTS reagent, followed by 1 h incubation at 37 °C in a 5% CO₂ incubator. After incubation, the 490 nm absorption of each well was measured by a light absorbance reader. Every group had 6 replicates.

Xenografts and animals. CWR22Rv1 stable cell lines with doxycycline-inducible shRNA against *lacZ*, *PRNCR1* or *PCGEM1* were injected into male athymic Nu/Nu mice, 4–5-weeks old. Mice arrived in our facility and were randomly put into cages with four mice each. Three million cells from each cell lines were suspended with Matrigel in the ratio of 2:1 for subcutaneous injection. The experiment was set up to use eight animals per group to detect a twofold increase in tumour size with power of 80% and at the significance level of 0.05 by a two-sided Student's *t*-test for significant studies. In these experiments, the tumour take rate was often about 50%, so we often ended up with about 4 animals per group, with or without doxycycline treatment. Tumour progression was monitored by caliper measurement twice a week and the tumour volume was calculated according to the equation, $v = \text{length} \times \text{width}^2 \times 1/2$. When the tumour size reached between 150–200 mm³, animals were randomly distributed to groups receiving 1 mg ml⁻¹ of doxycycline water or regular water with continuous tumour monitor until tumour burden exceeded the limit of IACUC humane endpoints (less than 20 mm in one dimension), for 3 to 4 weeks. The investigators were blinded to the group allocation during the experiment and/or when assessing the outcome. This study was carried out in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. Animals were housed in the UC Davis Medical Center Animal Facility under pathogen-free conditions, using a protocol approved by the Institutional Animal Care and Use Committee.

RNA isolation and qRT-PCR. Total RNA was isolated from cells using RNeasy Plus Mini Kit (Qiagen) following the manufacturer's protocol. First-strand cDNA synthesis from total RNA was carried out using iScript cDNA Synthesis Kit (Bio-Rad). Resulting cDNA was then analysed by quantitative PCR (qPCR) using Sso Advanced SYBR green supermix (Bio-Rad) on Stratagene Mx3000 or CFX Connect Real-Time PCR detection system. Primers were specific for genes tested and their sequences are listed in the oligonucleotide sequences and primers section. All qPCRs were repeated at least three times.

Oligonucleotide sequences (5'–3') and primers (forward and reverse). Antisense Oligonucleotide: Scrambled (mA*mA*mG*mC*mC*mC*G*C*A*C*C*A*G*C*G*mC*mC*mU*mC*mC)³⁰, *PRNCR1* #1 (mC*mC*mC*mU*mC*mC*T*C*C*T*T*C*T*mC*mU*mU*mG*mC), *PRNCR1* #2

(mA*mC*mU*mC*mU*C*T*T*T*C*T*T*C*A*C*mC*mU*mC*mC*mA), *PRNCR1* #3 (mA*mC*mU*mC*mC*C*A*C*A*C*A*C*A*mC*mC*mA*mC*mC), *PCGEM1* #1 (mU*mU*mC*mC*mC*T*T*C*T*G*C*T*T*G*C*C*mU*mG*mU*mU*mG), *PCGEM1* #2 (mG*mC*mU*mU*mU*A*C*C*C*T*T*A*G*T*C*mC*mU*mC*mC*mA) and *PCGEM1* #3 (mA*mG*mU*mC*mC*T*T*C*A*C*G*T*G*C*C*mU*mA*mC*mC*mC).

shRNA: *lacZ* (AAAAGCAGTTATCTGGAAGATCAGGTTGGATCCAACCTGATCTTCCAGATAACTGC) (Bioseitta); *PRNCR1* #1 (AAAAACAGTTTGATTA GGGAGGCACATTTATTGGATCCAATAAATGTGCCTCCCTAATCAAACCTG); *PRNCR1* #2 (AAAAAAGGAAGACTTTCCAGCACCTAATTGGATCCAATTA AGGTGCTGGAAAGTCCTTCCTT); *PCGEM1* #1 (AAAACCTTTGCAGAGAGC ATGCTTTCCTATTGGATCCAATAGGAAAGCATGCTCTCTGCAAAGG); *PCGEM1* #2 (AAAAACAACCTTTGCAGAGAGCATGCTTTTGGATCCAATAAAGC ATGCTCTCTGCAAAGGTTGT); and *AR* (AAAAGGTTCTCTGCTAGACGACA TTGGATCCAATGTCGTCTAGCAGAGAACC)¹³.

qPCR primer for gene expression and RIP: Actin (GCTCGTCGTCGACAACG GCTC and CAAACATGATCTGGGTCATCT); *PRNCR1* (CCAGATTCCAAGG CTGATA and GATGTTTGAGGCATCTGGT); *PCGEM1* (GGTGCTTTGCA ATGTTAT and AGCATGCTCTCTGCAAAGGT); *NEAT2* (TGGGGGAGT TTCTGACTGAG and TCTCCAGGACTTGCCAGTCT); *TMPRSS2* (CTGGTG GCTGATAGGGGAT and GTCTGCCCTCATTTGTCTGAT); *KLK2* (AGCCTGC CAAGATCACAGAT and GCAAGAAT CCTCTGGTTCTG); *PSA* (GATGCTG TGAAGGTCATGGA and TGGAGGTCCACACACTGAAAG); *FKBP5* (TCCCTC GAATGCAACTCTCT and GCCACATCTCTGCAGTCAAA); *NKX3-1* (GCCA AGAACCTCAAGCTCAC and AGAAGGCTCCTCTTTCAGG); *PGC-1* (GAG TTCCTGAGGACCCACAA and AGGAAGTTCTGGGGTGGAGT); *FASN* (AG GATCAGAGGACAACCTG and ACTCCACAGGTGGGAACAAG); and *NDRG1* (ACCTGCTACAACCCCTCTT and TGATCCATGGAGGGGTACAT).

qPCR primer for ChIRP and ChIP: *PSA* enhancer (TGGGACAACCTGCAAAC CTG and CCAGAGTAGGTCTGTTTCAATCCA); *PSA* middle region (CAGTG GCCATGAGTTTTGTTT and AACCAATCC AAC TGC ATT ATA CAC A); *PSA* promoter (CCTAGATGAAGTCTCCATGAGCTACA and GGGAGGGAGAGC TAGCACTTG); *KLK2* enhancer (GTTGAAAGCAGACCTACTCTGGA and GC ATATTTGTACAGCAGATAGCC); *KLK2* middle region (ATCTCAAGGACTT CTGGGTGGA and TGGGTAGTCCCTGTTACAAGAT); *KLK2* promoter (GG GAATGCCTCCAGACTGATC and CTTGCCCTGTTGGCACCTAGA); *TMPRSS2* enhancer (TGGTCTGGATGATAAAAAAAGTTT and GACATACGCCCCAC

AACAGA); *TMPRSS2* middle region (CCAGAAGAATACAATGATTAAAAGG CT and TGGAACTGAAGTATTGGAAAACCA); *TMPRSS2* promoter (CTGAG CCCCCACAATTGCAAAAC and GGTGGGACACACCTCAGCC); *RBL1* pro- moter (CAGCGTGGGGCTTGTCTCTCG and AGCGGAGGCAGACGGTGGAT)⁹; *MyoD1* enhancer (CAGCCAAGTATCCTCTCCA and AAGCTGAGCACTCT GGGAGA)¹⁴; *FASN* enhancer (CTACTTCTCCCGTGCCACTC and TCTCTCC CTTTCGATGTGTC); and *NDRG1* enhancer (GGTCACATCCAAGTGGGACT and AGAAGGTGGAGAGGGCAGTT).

PCR primer for *PSA* enhancer: (ATAGGGTTGGGCACTCACAGCTGAAT and AATGCTGGCAGAGTCCATGAGACTCC).

RT-PCR primer for AR variants: The full-length and truncated ARs were detec- ted using primers as described¹². F1: TGTCATATGGAGCTCTCACATGTGG; R1: CACCTCTCAAATATGCTAGACGAATCTGT; R2: GTACTCATTCAAGT ATCAGATATGCGGTATCAT; F6: CCATCTTGTCTGCTTTCGGAAATGTTA TGAAGC; R8: AGCTTCTGGGTTGTCTCTCTCAGTGG.

- Zhao, J. *et al.* Genome-wide identification of polycomb-associated RNAs by RIP- seq. *Mol. Cell* **40**, 939–953 (2010).
- Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
- Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
- Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
- Saeed, A. I. *et al.* TM4 microarray software suite. *Methods Enzymol.* **411**, 134–193 (2006).
- Zang, C. *et al.* A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25**, 1952–1958 (2009).
- Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
- Liu, W. *et al.* PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. *Nature* **466**, 508–512 (2010).
- Tsai, M. C. *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689–693 (2010).
- Liao, D. F., Monia, B., Dean, N. & Berk, B. C. Protein kinase C- ζ mediates angiotensin II activation of ERK1/2 in vascular smooth muscle cells. *J. Biol. Chem.* **272**, 6146–6150 (1997).

Mitotic trigger waves and the spatial coordination of the *Xenopus* cell cycle

Jeremy B. Chang¹ & James E. Ferrell Jr^{1,2}

Despite the large size of the *Xenopus laevis* egg (approximately 1.2 mm diameter), a fertilized egg rapidly proceeds through mitosis in a spatially coordinated fashion. Mitosis is initiated by a bistable system of regulatory proteins centred on Cdk1 (refs 1, 2), raising the possibility that this spatial coordination could be achieved through trigger waves of Cdk1 activity³. Using an extract system that performs cell cycles *in vitro*, here we show that mitosis does spread through *Xenopus* cytoplasm via trigger waves, propagating at a linear speed of approximately $60 \mu\text{m min}^{-1}$. Perturbing the feedback loops that give rise to the bistability of Cdk1 changes the speed and dynamics of the waves. Time-lapse imaging of intact eggs argues that trigger waves of Cdk1 activation are responsible for surface contraction waves, ripples in the cell cortex that precede cytokinesis^{4,5}. These findings indicate that Cdk1 trigger waves help ensure the spatiotemporal coordination of mitosis in large eggs. Trigger waves may be an important general mechanism for coordinating biochemical events over large distances.

During mitosis, a cell undergoes marked, abrupt reorganization: the chromosomes condense, the Golgi and endoplasmic reticulum fragment, and the dynamics of the microtubules changes. These processes are driven by the activation of the protein kinase Cdk1. In a typical somatic cell, with a cellular radius of the order of $10 \mu\text{m}$, it should be relatively easy to achieve a near-synchronous onset of mitosis in all parts of the cell. If Cdk1 were activated abruptly in the middle of the cell (current evidence suggests that Cdk1 is first activated on the centrosomes⁶), then it should take no more than a few seconds for the active complexes to spread by random walk diffusion to all parts of the cell ($t = \langle d^2 \rangle / 6D$; assuming the diffusion coefficient $D = 10 \mu\text{m}^2 \text{s}^{-1}$, $t \approx 2 \text{s}$). This would be sufficiently fast to account for the observed near-synchronicity of mitotic entry in somatic cells.

However, some eukaryotic cells, notably the well-studied *X. laevis* egg, are much larger; a *Xenopus* egg is approximately $600 \mu\text{m}$ in radius, which corresponds to a diffusion time on the order of 2 h rather than 2 s. This suggests that mitosis might be a sluggish process in these cells.

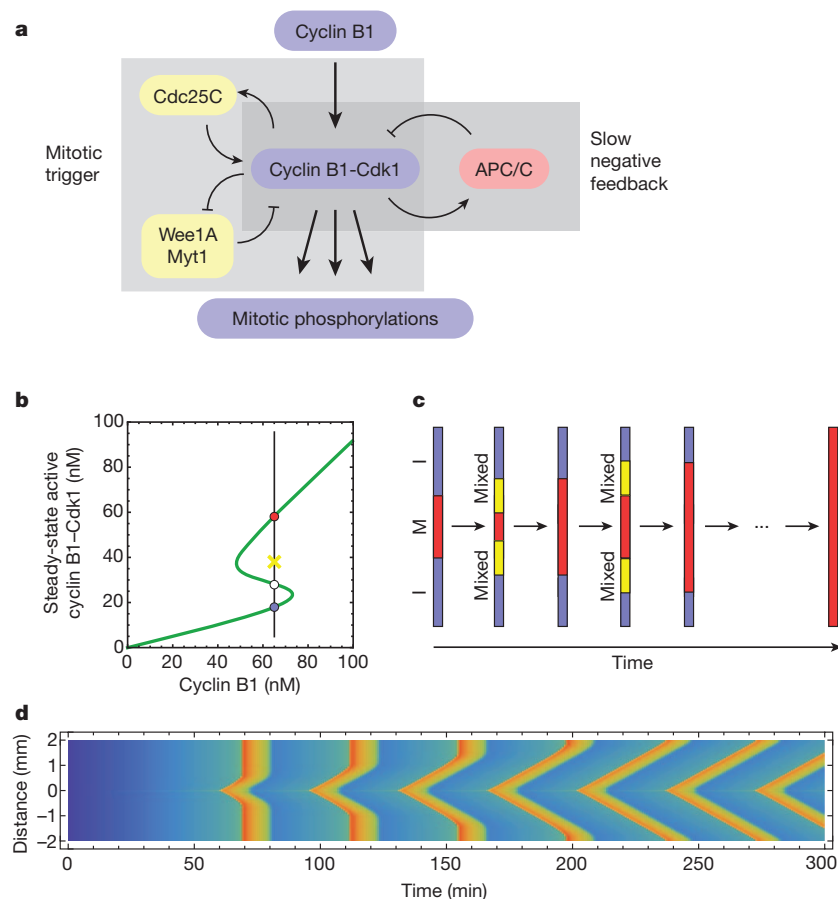


Figure 1 | Trigger waves in Cdk1 activation.

a, Schematic view of the Cdk1-APC/C circuit. **b**, Modelled steady-state response of Cdk1 to cyclin B1, with parameters on the basis of experimental studies^{25,26}. At intermediate concentrations of cyclin, the response is bistable. The stable low (interphase) and high (mitotic) Cdk1 activity steady-states corresponding to one such cyclin concentration (65 nM) are shown as blue and red circles, respectively. The white circle denotes an unstable steady state, and the yellow X denotes an intermediate level of Cdk1 activity that would be attracted to the mitotic (red) steady state. **c**, Schematic view of the propagation of Cdk1 activity up and down a one-dimensional tube through successive rounds of mixing and conversion. I, interphase; M, mitosis. **d**, Modelled propagation of Cdk1 activity in a one-dimensional tube. It is assumed that cyclin B1 is synthesized at a uniform rate everywhere in the tube, but in a $5 \mu\text{m}$ region in the middle of the tube the concentration of Cdc25C is 50% higher than in the rest of the tube, allowing Cdk1 to become activated earlier. Cyclin B1-Cdk1 activity is denoted by the colour scale (blue is low, red is high). Numerical solution of the partial differential equations used Mathematica 9.0 (Wolfram) as described in the Supplementary Materials. See also Supplementary Fig. 1.

¹Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, California 94305-5174, USA. ²Department of Biochemistry, Stanford University School of Medicine, Stanford, California 94305-5307, USA.

Actually, the cell cycle is even more rapid in early *Xenopus* embryos than in somatic cells⁷. Moreover, mitotic events occur within minutes of each other in distant parts of the cell. For example, nuclear envelope breakdown, which occurs in the interior of the fertilized egg, takes place only a few minutes before the mitotic surface contraction waves (SCWs; discussed in more detail below) sweep over the cortex of the egg⁸. Thus, some mechanism other than simple diffusion of active Cdk1 must coordinate spatially distant mitotic events.

One possible mechanism is suggested by the systems-level organization of the mitotic trigger circuit, which includes interlinked positive and double-negative feedback loops (Fig. 1a). Circuits such as this can show bistability; indeed, experimental studies in *Xenopus* egg extracts have shown that the response of Cdk1 to non-degradable cyclin B1 is hysteretic, with two alternative stable levels of Cdk1 activity possible at intermediate cyclin B1 concentrations (Fig. 1b)^{1,2}. Bistability could, in principle, allow Cdk1 activation to propagate rapidly through trigger waves^{3,9–11}.

To see why bistability can generate rapidly propagating trigger waves, imagine a long thin tube containing cytoplasm with a uniform concentration of cyclin B1 and Cdk1, and assume that in some region of the tube the cytoplasm is in the mitotic, high Cdk1-activity state (Fig. 1c, red) while the rest of the cytoplasm is in the interphase, low Cdk1-activity state (Fig. 1c, blue). Within some small distance of the interface, the cytoplasm will rapidly mix by diffusion, resulting in an intermediate level of Cdk1 activity. If this activity is above the unstable steady state (Fig. 1b, white point), this slice of cytoplasm will flip to the mitotic state. The process of mixing and conversion repeats, and the mitotic state propagates down the tube at a constant velocity. The propagation speed can be estimated by Luther's formula, $v \approx 2(D/\tau)^{1/2}$, where D is the diffusion coefficient and τ is related to the flipping time

for the bistable system^{3,12}. If we assume $D = 10 \mu\text{m}^2 \text{s}^{-1}$ and $\tau = 10\text{--}100 \text{ s}$, the expected propagation speed would be $40\text{--}120 \mu\text{m min}^{-1}$, and Cdk1 activity could propagate from the congressed pronuclei to the animal pole in 2–5 min. This would be fast enough to account for the coordination of nuclear and cortical mitotic events. If Cdk1 activation is followed by APC/C activation and cyclin degradation, a wave of mitotic exit would follow the wave of mitotic entry, and if this is followed by new cyclin synthesis the whole process would repeat.

To explore this idea further, we created a simple partial differential equation (PDE) model of Cdk1 activation and propagation (Supplementary Materials). We assumed that cyclin B1 was synthesized at a uniform, constant rate everywhere in the cytoplasm, and that the mitotic activator Cdc25C was 50% higher in concentration in one 5 μm section of the tube than in the rest of it. This inhomogeneity could represent the centrosome, which, in somatic cells, has a high concentration of Cdc25C¹³. As anticipated, the modelled activation of Cdk1 was found to occur first in this high Cdc25C region, and then to spread linearly up and down the tube, resulting in a V-shaped front of Cdk1 activation in the plot of activity as a function of time and position (Fig. 1d; see also Supplementary Fig. 1). The propagation rate, which is the slope of the diagonal wave fronts, was approximately $60 \mu\text{m min}^{-1}$, compatible with the estimates from Luther's formula. Farther from the centrosomal region, cyclin synthesis reached the threshold for Cdk1 activation before the trigger wave arrived, resulting in a vertical front of Cdk1 activity. However, with successive cycles, the trigger waves came to occupy more and more of the tube. These results support the plausibility of trigger waves as a mechanism for allowing Cdk1 activation to spread through an egg in a rapid and orderly manner.

We therefore set out to determine experimentally whether trigger waves do occur, using cycling *X. laevis* egg extracts¹⁴. De-membranated

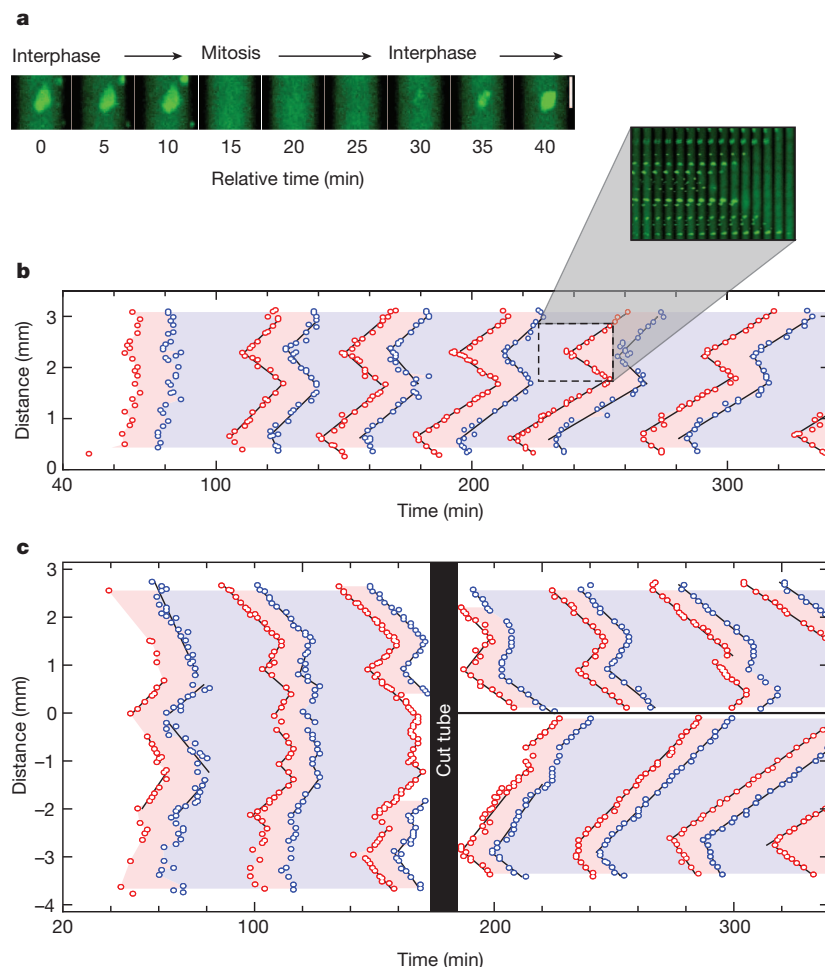
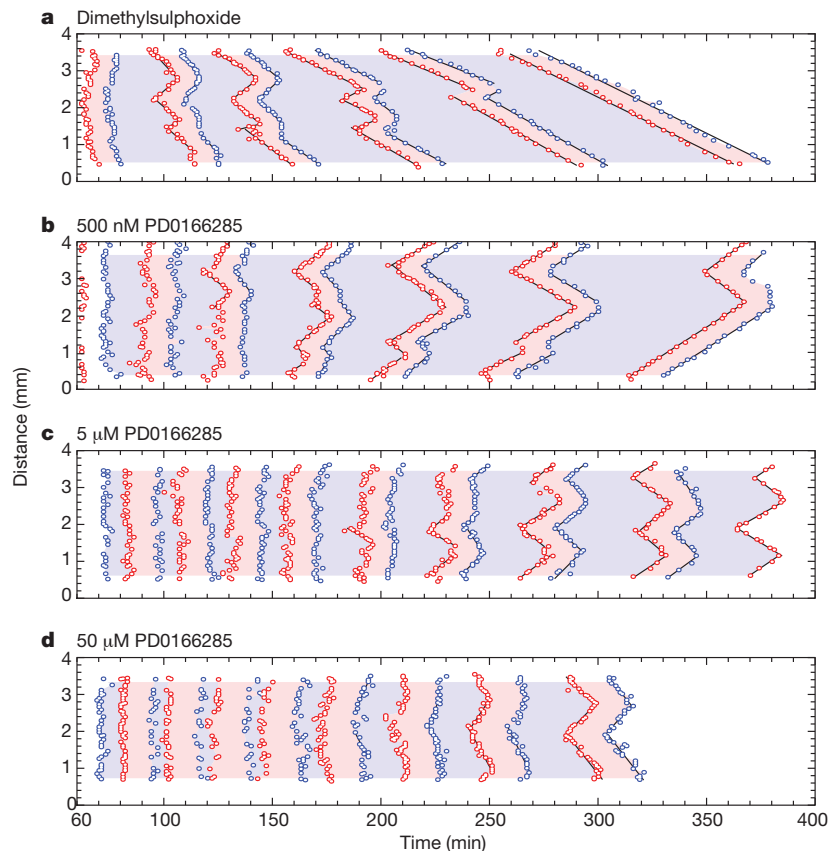


Figure 2 | Rapid, linear propagation of mitotic entry and exit through *Xenopus* cytoplasm. **a**, An example of nuclear envelope breakdown and nuclear envelope reformation in an extract with added sperm chromatin and GFP-NLS. **b**, The timing of mitotic entrance and exit in an approximately 3 mm length of Teflon tubing submerged in mineral oil. Each data point represents the time and position at which an individual nucleus underwent nuclear envelope breakdown (red points) or nuclear envelope reformation (blue points). The pink and blue regions of the plot denote mitosis and interphase, respectively. Time is measured relative to when the extract was warmed to room temperature. The inset shows frames from the video in montage form. **c**, Trigger waves versus phase waves. The tube was cut under mineral oil at 160 min. See also Supplementary Fig. 2.

Figure 3 | The Wee1/Myt1 inhibitor PD0166285 accelerates the trigger waves. a–d, Mitotic entrance and exit waves in extracts treated with dimethylsulphoxide or one of three concentrations of PD0166285. See also Supplementary Fig. 3.



sperm chromatin and purified green fluorescent protein (GFP)–nuclear localization signal (NLS) protein were added to the extracts as reporters of mitotic entry. In interphase extracts, the sperm chromatin forms nuclei, complete with nuclear pores that import the GFP–NLS protein. As the cell cycle progresses and the extract enters mitosis, the nuclear envelopes break down and GFP–NLS quickly disperses (Fig. 2a). About 15 min later the nuclei begin to reform (Fig. 2a), indicating the end of mitosis. We loaded extract into a Teflon tube, submerged the tube in mineral oil, and cut the tube into sections 2–3 mm long. These sections were then imaged by time-lapse epifluorescence microscopy. Care was taken to keep the temperature of the tube constant to avoid convective flow. Raw microscopy images were stitched together, and the times and locations of nuclear envelope breakdown and re-formation were recorded.

Supplementary Video 1 and Fig. 2b show the behaviour of a typical cycling extract. The nuclei in the extract initially entered mitosis about 60 min after the extract was warmed to room temperature, and then continued to undergo cycles of interphase and mitosis approximately every 40 min over the next several hours. During the first cycle, the mitoses occurred nearly synchronously, with some apparently stochastic variation in timing from nucleus to nucleus. From the second cycle on, both mitotic entry and mitotic exit swept through the tube in a linear, wave-like fashion (Fig. 2b), with the waves initiating near the same locations during every cycle. Initially the wave speed was $54 \pm 20 \mu\text{m min}^{-1}$ (mean \pm s.d.), in good agreement with the expected speeds of trigger waves on the basis of either Luther's formula or our PDE simulations (Fig. 1). The waves gradually slowed to about half that speed by the end of the experiment (Fig. 2b and Supplementary Fig. 2). Generally the waves became better organized and spanned longer lengths of the channel as the experiment proceeded (Figs 2b, c and 3a). These findings suggest that waves of mitosis emanating from the earliest or strongest foci eventually dominate the dynamics of the whole system.

In principle, the observed waves could actually be phase waves (also called pseudowaves or kinematic waves^{3,15}), which could occur if different

parts of the cytoplasm initiated oscillations at different times and this temporal difference was then faithfully maintained in the subsequent waves. To test this possibility, we placed a tube of extract in mineral oil, monitored three cycles of mitotic entry and exit, and then cut the tube into two pieces with a scalpel. As shown in Fig. 2c and Supplementary Video 2, the nuclei on either side of the cut quickly lost coordination, and seemed to follow the rhythm established by the fastest-initiating region of cytoplasm on their side of the cut. This eliminates the possibility that the observed waves are simply phase waves.

If the Cdk1 network generates trigger waves, then changing the dynamics of the feedback loops that regulate Cdk1 should affect wave dynamics. To test this idea, we made use of PD0166285, a small molecule inhibitor of Wee1A and Myt1 (ref. 16). Experiments in bulk extracts showed that inhibition of cyclin B1-induced Cdk1 tyrosine phosphorylation was half-maximal at approximately 1 μM (Supplementary Fig. 3). We treated extracts with 0.5, 5 or 50 μM PD0166285 or just dimethylsulphoxide, and monitored the effects on the timing of mitotic onset and the speed of mitotic propagation.

As shown in Fig. 3 and Supplementary Fig. 3, PD0166285 advanced the onset of mitosis by approximately 30 min and increased the speed of the mitotic waves to as much as $120 \mu\text{m min}^{-1}$. Interphase durations were shortened by as much as 50%, and there was an increase in mitotic duration (Supplementary Fig. 3), which emphasizes the importance of Wee1A/Myt1 activation in mitotic exit. At the highest concentrations of PD0166285, waves were not evident until late in the experiment (Fig. 3d). These findings support the hypothesis that the waves of mitotic entry and exit arise from the bistable mitotic trigger.

Finally, we turned to the question of whether the trigger waves seen in extracts also occur in intact fertilized eggs. We considered that waves of Cdk1 activation could be initiated at the centrosome, spread outwards through the cyclin B1–Cdk1 complexes diffusely present throughout the interphase cytoplasm¹⁷, and then cause SCWs when they reach the cortical cytoskeleton. SCWs are waves of pigmentation change at the cell cortex that proceed from the animal pole to the vegetal pole just

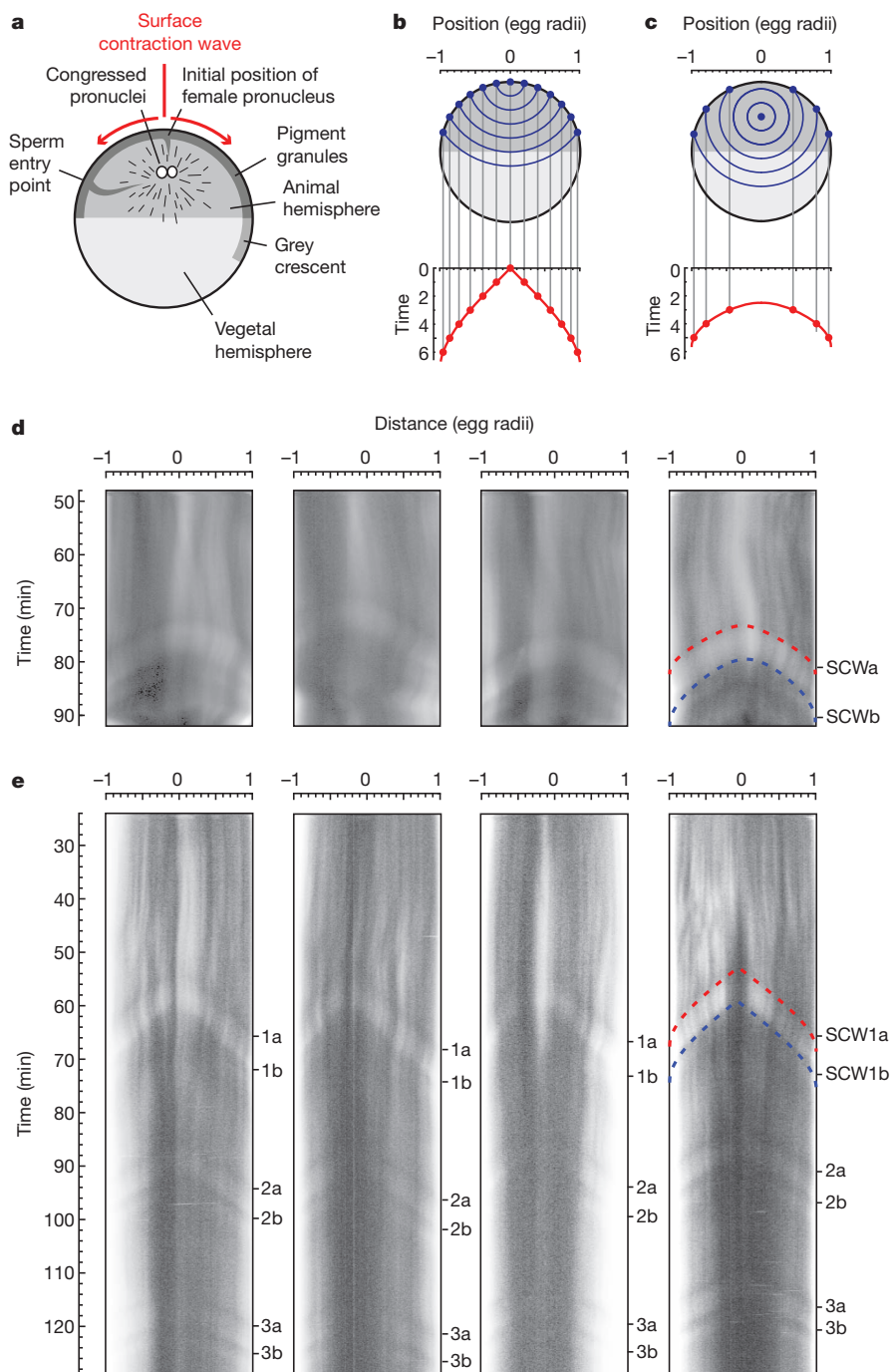


Figure 4 | Surface contraction waves in intact *Xenopus* eggs. **a**, Schematic view of the anatomy of a fertilized egg just before the onset of mitosis (adapted from ref. 27). **b**, **c**, Expected propagation of surface contraction waves if they were due to a spherical wave of Cdk1 activation spreading from a point source. An equation for the spreading of the waves (Supplementary Equation 1) is derived in the Supplementary Materials. Panel **b** assumes the point source is at the animal pole. Panel **c** assumes the point source is halfway between the animal pole and the centre of the cell. **d**, **e**, Kymographs depicting surface contraction waves, indicated by transitions from light to dark or dark to light, in four fertilized eggs (**d**) and four parthenogenetically activated eggs (**e**). The red and blue dashed curves are fits of the experimental data to Supplementary Equation 1. See also Supplementary Fig. 4.

before cytokinesis^{4,18,19} (Supplementary Video 3 and Fig. 4a). The first wave (SCWa) is linked to Cdk1 activation, and the second wave (SCWb) requires Cdk1 inactivation¹⁸. It has been speculated that the change in pigmentation is controlled by the phosphorylation of microtubule associated proteins by Cdk1 (ref. 20).

We derived a formula for the time at which such a constant velocity, spherical wave would reach various points on the surface of the egg as a function of the position of origin of the wave and the wave speed (Supplementary Materials, Supplementary Equation 1 and Supplementary Fig. 4). If the trigger wave were to begin in the middle of the cell, we would expect it to arrive at all points on the cortex simultaneously. If it were to originate from the animal pole, we would expect a plot of the wave's cortical position as a function of time to be V-shaped (Fig. 4b). If it originated in the middle of the animal hemisphere, we would expect the plot to be U-shaped (Fig. 4c).

We then imaged SCWs travelling across the animal hemispheres of fertilized eggs and expressed the data as kymographs (Fig. 4d). The position versus time curves were U-shaped and agreed well with Supplementary Equation 1 (Fig. 4d). By fitting the observed SCWa curves to Supplementary Equation 1, we inferred a trigger wave speed of $66 \pm 17 \mu\text{m min}^{-1}$ (mean \pm s.d., $n = 4$). This speed is similar to that observed in extracts (Figs 2 and 3). The inferred origin of the waves was at a location $85 \pm 5\%$ (mean \pm s.d., $n = 4$) up from the vegetal pole on the animal-vegetal axis. This origin is near where the congressed pronuclei reside (Fig. 4a), and the 99% confidence intervals for the location of the origin in each embryo (87–93%, 76–91%, 71–85% and 75–88%) essentially rule out the possibility that the trigger waves originated at the animal pole. The inferred time at which the mitotic trigger wave originated was approximately 70 min after fertilization and 3 ± 1 min (mean \pm s.d., $n = 3$) before the beginning of SCWa.

Parthenogenetically activated eggs do not undergo cleavages, but do show a succession of SCWs⁵. Because pronuclear congression does not occur in parthenogenetically activated eggs, it seemed plausible that the spherical mitotic trigger waves considered to give rise to these SCWs might have a different point of origination. As shown in Fig. 4e and Supplementary Video 4, parthenogenetically activated eggs showed a very regular succession of SCWs whose position versus time curves were V-shaped rather than U-shaped. Fitting these curves to Supplementary Equation 1 yielded an inferred wave speed of $58 \pm 3 \mu\text{m min}^{-1}$ (mean \pm s.d., $n = 4$) and a point of origination immediately under the animal pole. The 99% confidence intervals for the locations were 99–101%, 98–100%, 99–102% and 98–102% up from the vegetal pole. Thus, the quantitative characteristics of the SCWs both in fertilized and in parthenogenetically activated eggs are consistent with the hypothesis that they occur when trigger waves of Cdk1 activation and inactivation arrive at the cell cortex.

In summary, we have shown through modelling studies that the Cdk1/Wee1A/Cdc25C system could plausibly generate trigger waves of sufficient speed to help spatially coordinate mitosis in the large, fertilized *X. laevis* egg, and have shown experimentally that linear waves of mitotic entry and exit do occur in unstirred cycling *Xenopus* egg extracts and are likely to occur in intact eggs. Note that the early cell cycles of the syncytial *Drosophila* embryo also show mitotic waves. By cycles 10–13, mitosis can be seen to begin at the two poles and spread towards the middle of the embryo, at a rate of approximately $50\text{--}250 \mu\text{m min}^{-1}$ (ref. 21). Mitotic waves have also been detected in some multinucleate fungi (for example *Stemonitis flavogenita*²² and *Aspergillus nidulans*²³, but not in *Ashbya gossypii*²⁴). We suspect that all of these mitotic waves are essentially the same phenomenon we are observing here.

Random walk diffusion is rapid enough to allow proteins to communicate between different parts of a small cell on a timescale of seconds, and communication in large organisms often takes place by flow. Trigger waves provide another mechanism for communicating rapidly over millimetre to metre distance scales. Trigger waves are the basis for the propagation of action potentials down an axon, for the propagation of calcium waves through excitable cells and tissues, and for cyclic AMP waves in *Dictyostelium discoideum*⁹. Mitotic waves seem to be another example of this phenomenon. The proteins, biochemical processes and timescales involved in action potentials and mitotic waves are different, but the basic systems-level logic is the same. This suggests that trigger waves may be an important general mechanism for communication within tissues, organs and organisms.

METHODS SUMMARY

Time-lapse epifluorescence microscopy of *Xenopus* cycling extracts. Cycling extracts were prepared¹⁴ and, while kept on ice, supplemented with de-membranated sperm chromatin (to approximately 400 per microlitre of extract) and GFP–NLS (to approximately 10 μM). Extracts were then taken off ice and a portion was loaded into Teflon tubing (Cole-Parmer, number YO-06417-72), submerged in mineral oil on a glass slide and cut into 2–3 mm sections.

The tube was imaged at room temperature (22–25 °C) on an inverted epifluorescence microscope (Leica DMI6000 B). Samples from extract not loaded into the tubes were frozen down and later assayed for histone H1 kinase activity.

For studies using PD0166285, extracts were supplemented with a fixed volume of dimethylsulphoxide or inhibitor along with the sperm chromatin and GFP–NLS. **Image analysis.** Images of the channels were cropped, stitched together and contrast-adjusted using a custom MATLAB (Mathworks) script. The locations and times at which nuclei underwent envelope breakdown or formation were recorded manually.

Received 28 March; accepted 22 May 2013.

Published online 17 July 2012.

1. Pomerening, J. R., Sontag, E. D. & Ferrell, J. E. Jr. Building a cell cycle oscillator: hysteresis and bistability in the activation of Cdc2. *Nature Cell Biol.* **5**, 346–351 (2003).

2. Sha, W. *et al.* Hysteresis drives cell-cycle transitions in *Xenopus laevis* egg extracts. *Proc. Natl Acad. Sci. USA* **100**, 975–980 (2003).
3. Novak, B. & Tyson, J. J. Modeling the cell division cycle: M-phase trigger, oscillations, and size control. *J. Theor. Biol.* **165**, 101–134 (1993).
4. Hara, K. Cinematographic observation of “surface contraction waves” (SCW) during the early cleavage of axolotl eggs. *Wilhelm Roux Arch. Dev. Biol.* **167**, 183–186 (1971).
5. Hara, K., Tydemann, P. & Kirschner, M. A cytoplasmic clock with the same period as the division cycle in *Xenopus* eggs. *Proc. Natl Acad. Sci. USA* **77**, 462–466 (1980).
6. Jackman, M., Lindon, C., Nigg, E. A. & Pines, J. Active cyclin B1–Cdk1 first appears on centrosomes in prophase. *Nature Cell Biol.* **5**, 143–148 (2003).
7. Newport, J. & Kirschner, M. A major developmental transition in early *Xenopus* embryos: I. Characterization and timing of cellular changes at the midblastula stage. *Cell* **30**, 675–686 (1982).
8. Gerhart, J. C. in *Biological Regulation and Development* Vol. 2 (ed. Goldberger, R. F.) 133–316 (Plenum, 1980).
9. Tyson, J. J. & Keener, J. P. Singular perturbation theory of traveling waves in excitable media (a review). *Physica D* **32**, 327–361 (1988).
10. Markevich, N. I., Tsyganov, M. A., Hoek, J. B. & Kholodenko, B. N. Long-range signaling by phosphoprotein waves arising from bistability in protein kinase cascades. *Mol. Syst. Biol.* **2**, 61 (2006).
11. Reynolds, A. R., Tischer, C., Verveer, P. J., Rocks, O. & Bastiaens, P. I. EGFR activation coupled to inhibition of tyrosine phosphatases causes lateral signal propagation. *Nature Cell Biol.* **5**, 447–453 (2003).
12. Luther, R. Räumliche Fortpflanzung chemischer Reaktionen. *Z. Elektrochemie* **12**, 596–600 (1906).
13. Bonnet, J., Coopman, P. & Morris, M. C. Characterization of centrosomal localization and dynamics of Cdc25C phosphatase in mitosis. *Cell Cycle* **7**, 1991–1998 (2008).
14. Murray, A. W. Cell cycle extracts. *Methods Cell Biol.* **36**, 581–605 (1991).
15. Winfree, A. T. Two kinds of wave in an oscillating chemical solution. *Faraday Symp. Chem. Soc.* **9**, 38–46 (1974).
16. Wang, Y. *et al.* Radiosensitization of p53 mutant cells by PD0166285, a novel G(2) checkpoint abrogator. *Cancer Res.* **61**, 8211–8217 (2001).
17. Nakamura, N., Tokumoto, T., Ueno, S. & Iwao, Y. The cytoskeleton-dependent localization of cdc2/cyclin B in blastomere cortex during *Xenopus* embryonic cell cycle. *Mol. Reprod. Dev.* **72**, 336–345 (2005).
18. Rankin, S. & Kirschner, M. W. The surface contraction waves of *Xenopus* eggs reflect the metachronous cell-cycle state of the cytoplasm. *Curr. Biol.* **7**, 451–454 (1997).
19. Shinagawa, A., Konno, S., Yoshimoto, Y. & Hiramoto, Y. Nuclear involvement in localization of the initiation site of surface contraction waves in *Xenopus* eggs. *Dev. Growth Differ.* **31**, 249–255 (1989).
20. Perez-Mongiovi, D., Chang, P. & Houliston, E. A propagated wave of MPF activation accompanies surface contraction waves at first mitosis in *Xenopus*. *J. Cell Sci.* **111**, 385–393 (1998).
21. Foe, V. E. & Alberts, B. M. Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in *Drosophila* embryogenesis. *J. Cell Sci.* **61**, 31–70 (1983).
22. Haskins, E. F. *Stemonitis flavogenita* (Myxomycetes) – plasmodial phase (aphanoplasmodium). *Film E2000, Institut Wissenschaftliche Film, Göttingen. Pub. Wiss. Film Sect. Biol.* **8**, 1–14 (1974).
23. Clutterbuck, A. J. Synchronous nuclear division and septation in *Aspergillus nidulans*. *J. Gen. Microbiol.* **60**, 133–135 (1970).
24. Gladfelter, A. S., Hungerbuehler, A. K. & Philippsen, P. Asynchronous nuclear division cycles in multinucleated cells. *J. Cell Biol.* **172**, 347–362 (2006).
25. Trunnell, N. B., Poon, A. C., Kim, S. Y. & Ferrell, J. E., Jr. Ultrasensitivity in the regulation of Cdc25C by Cdk1. *Mol. Cell* **41**, 263–274 (2011).
26. Yang, Q. & Ferrell, J. E., Jr. The Cdk1–APC/C cell cycle oscillator circuit functions as a time-delayed ultrasensitive switch. *Nature Cell Biol.* **15**, 519–525 (2013).
27. Hausen, P. & Riebesell, M. in *The Early Development of Xenopus laevis: An Atlas of the Histology* plate 12 (Springer, 1991).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank H. Funabiki and M. Dasso for providing GFP–NLS protein and constructs, E. Sontag and L. Wang for discussions and calculations, T. Tsai for sharing his unpublished data on the effects of PD0166285 on *Xenopus* embryos, Q. Yang for helping to build the ordinary differential equation model upon which our partial differential equation model is based, G. Dey, H. Stone and B. Sullivan for discussions, S. Quake and the Quake laboratory for advice, J. Chen and the Chen laboratory for the use of their microscope and discussions, J. Pomerening for discussions and technical advice, the Stanford Cell Sciences Imaging Facility for technical assistance, and members of the Ferrell laboratory for discussions. This work was supported by grants from the National Institutes of Health (GM046383 and GM077544) and by a National Science Foundation Graduate Research Fellowship and a Lieberman Fellowship (to J.B.C.).

Author Contributions J.B.C. performed experiments and calculations, analysed data and helped write the paper. J.E.F. performed calculations, analysed data and helped write the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.B.C. (jbchang@stanford.edu) or J.E.F. (james.ferrell@stanford.edu).

Avoiding chromosome pathology when replication forks collide

Christian J. Rudolph^{1,2*}, Amy L. Upton^{1,2†}, Anna Stockum^{1†}, Conrad A. Nieduszynski¹ & Robert G. Lloyd^{1*}

Chromosome duplication normally initiates through the assembly of replication fork complexes at defined origins^{1,2}. DNA synthesis by any one fork is thought to cease when it meets another travelling in the opposite direction, at which stage the replication machinery may simply dissociate before the nascent strands are finally ligated. But what actually happens is not clear. Here we present evidence consistent with the idea that every fork collision has the potential to threaten genomic integrity. In *Escherichia coli* this threat is kept at bay by RecG DNA translocase³ and by single-strand DNA exonucleases. Without RecG, replication initiates where forks meet through a replisome assembly mechanism normally associated with fork repair, replication restart and recombination^{4,5}, establishing new forks with the potential to sustain cell growth and division without an active origin. This potential is realized when roadblocks to fork progression are reduced or eliminated. It relies on the chromosome being circular, reinforcing the idea that replication initiation is triggered repeatedly by fork collision. The results reported raise the question of whether replication fork collisions have pathogenic potential for organisms that exploit several origins to replicate each chromosome.

In *E. coli*, the number of head-on fork collisions is kept to a single event by replicating the circular chromosome from a single origin (*oriC*). Chromosome duplication is completed when the two forks established meet in a specialized termination zone (Fig. 1a). This zone is flanked by *ter* sequences bound by Tus protein, forming polar traps that restrict fork movement¹. Highly expressed genes are transcribed co-directionally with replication, minimizing the negative impact of conflict between transcription and replication^{6–8}.

Initiation at *oriC* is controlled by DnaA⁹. However, replication can initiate independently of both DnaA and *oriC*. This stable DNA replication is robust enough to sustain growth in strains lacking RNase HI, especially in minimal media¹⁰ (Supplementary Fig. 1). Stable DNA replication is also elevated in strains lacking RecG³, but in this case appears unable to sustain growth (Fig. 1b)¹⁰. Forks must either be established in too small a fraction of cells or they cannot proceed far enough to complete chromosome duplication. Replication profiling of *recG* cells by deep sequencing suggested the latter. Exponentially growing wild-type cells showed a clearly defined origin and termination region (Fig. 1c and Supplementary Figs 4 and 8)¹¹. The profile of *recG* cells, on the other hand, showed a high incidence of extra initiation at a sharply focused site in the terminus area (Figs 1c–d, 2d and 3c). Initiation at this site is exceptionally frequent in the genetic background of strain AB1157; it approaches that at *oriC* (Supplementary Fig. 2a). Replication is evident in both directions, but limited to amplification of the *terA–terB* interval. However, a combination of two mutations, one (*Δtus*) eliminating Tus/*ter* traps, and the other (*rpoB**35) destabilizing transcription complexes^{12,13}, enables replication to extend further, leading to a broader amplification (Fig. 1c). Notably, these two mutations together enable *recG* cells to grow

without DnaA or *oriC* (Fig. 1b and Supplementary Figs 1 and 3), demonstrating that stable DNA replication is now able to duplicate the entire chromosome. Viability assays on 17 independent exponential-phase cultures showed that $58 \pm 17\%$ of *dnaA recG tus rpo** cells detected by plating on LB agar at 30 °C establish colonies at 42 °C, which is a much higher viability than we see with *dnaA rnhA* cells (Fig. 1b). The colonies are small, but continue growing (Fig. 1b and Supplementary Fig. 4). The *oriC* deletion derivative shows similar slow growth, with a doubling time of 100 min as opposed to 21 min for the

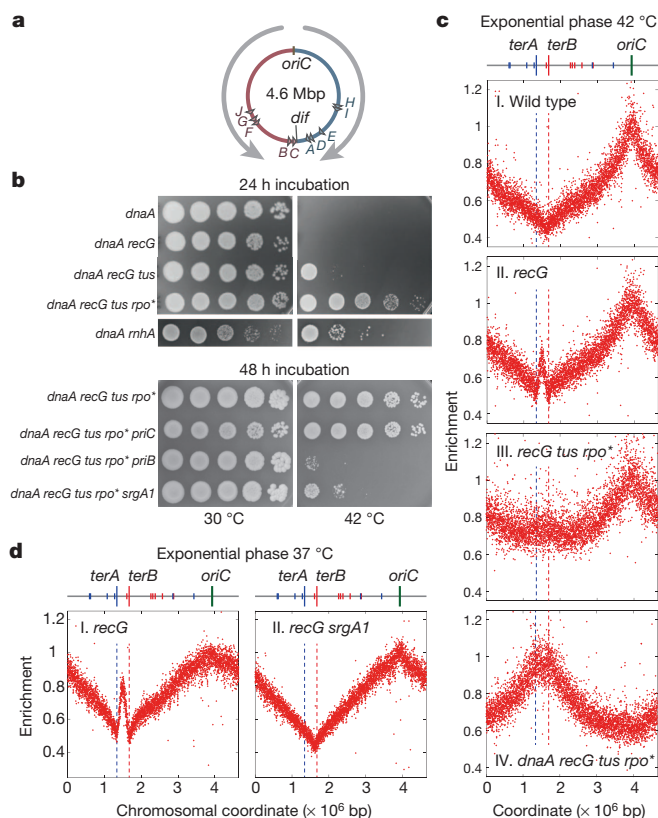


Figure 1 | PriA triggers DnaA-independent chromosome replication in the absence of RecG. **a**, Replicore arrangement of the *E. coli* chromosome. Grey arrows indicate the normal direction of replication and polarity of major transcription. Triangles indicate *ter* sites. **b**, Genetic analysis of growth without DnaA (*dnaA46* at 42 °C). The strains were AU1054, AU1091, RCe205, RCe268, AU1066, RCe268, N8201, N8205 and N8206. **c**, **d**, Replication profiling of exponential-phase cells. Normalized numbers of reads are plotted against chromosomal coordinate. Sequencing templates for **c** were from MG1655, N8226, RCe261 and RCe268, all cultured at 42 °C, and for **d** were from N6576 and JJ1268 cultured at 37 °C.

¹Centre for Genetics and Genomics, University of Nottingham, Queen's Medical Centre, Nottingham NG7 2UH, UK. ²Division of Biosciences, School of Health Sciences and Social Care, Brunel University, Uxbridge, London UB8 3PH, UK. [†]Present addresses: Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK (A.L.U.); Division of Medicine, Imperial College London, St Mary's Campus, Norfolk Place, London W2 1PG, UK (A.S.).

*These authors contributed equally to this work.

wild type. Marker frequency analysis showed an inverted profile for *dnaA recG tus rpo** cells at 42 °C, with a peak at the terminus and no evidence of initiation at *oriC* (Fig. 1c).

Replication initiated in the terminus area in *recG* cells is abolished by mutations that modulate PriA helicase activity (*srgA1*, *priA300*) or which eliminate PriB (Fig. 1d and Supplementary Fig. 4). Both proteins facilitate DnaB loading and replisome assembly at branched DNAs⁴. The same mutations prevent *dnaA recG tus rpo** cells from establishing colonies at 42 °C. Eliminating the restart protein PriC does not (Fig. 1b). 5-Bromo-2'-deoxyuridine (BrdU)-labelling confirmed that *priA300* reduces stable DNA replication (Supplementary Fig. 4b). Thus the DNA replication initiated at the terminus in cells lacking RecG must be the result of a PriA–PriB-mediated loading of DnaB at a branched DNA structure generated in this region. The ability of *srgA1* to prevent initiation (Fig. 1d) indicates that the branched DNA is a 3' flap. This allele encodes a PriA protein that has specifically lost the ability to unwind such a structure¹⁴. So 3' flaps might normally be unwound by RecG³ and eliminated by single-stranded DNA (ssDNA) exonucleases, reducing the likelihood of PriA targeting the structure and triggering replication.

This fits with the fact that 3' ssDNA exonuclease activity is needed to keep *recG* cells alive, and that *priA300* overcomes this requirement¹⁵. We profiled replication in exponential-phase cells lacking ExoI, ExoVII and SbcCD and observed that replication initiates with high frequency in the terminus area despite the presence of RecG (Fig. 2a). This strengthens the notion that replication initiates at 3' flaps. More importantly, it demonstrates that the re-replication observed without RecG is not a peculiarity of *recG* cells. Rather, the generation of branched DNA structures with a capacity to trigger replication appears

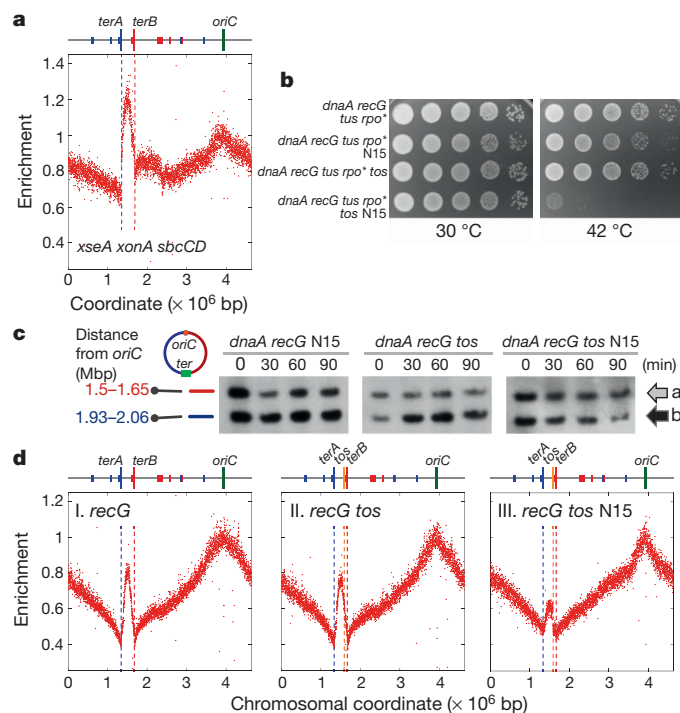


Figure 2 | Replisome collision triggers DnaA-independent replication. **a**, Replication profile of exonuclease-depleted cells. Sequencing templates were from exponential-phase N6953 cultured at 37 °C. **b**, DnaA-independent growth. The strains analysed were RCe268, RCe384, RCe385 and RCe387. **c**, BrdU labelling of a fragment of NotI-digested DNA (**b**) located near the replication terminus relative to a distant reference fragment (**a**). The strains were RCe405, RCe409 and RCe418. Cultures of *dnaA46* strains of the genotypes indicated were pulse labelled with BrdU at the indicated times after the shift to 42 °C. **d**, Replication profiling showing effect of chromosome linearization on replication. Sequencing templates were from N8226, RCe391 and RCe399 cultured at 37 °C.

to be a regular feature of the cell cycle triggered each time forks collide to complete replication.

We investigated what happens when forks are prevented from colliding by using strains in which the chromosome is linearized by phage N15 telomerase at a *tos* site inserted near *dif*⁶ (Supplementary Fig. 5). Linearization has no effect on the growth of wild-type cells (Supplementary Fig. 5h)¹⁶. However, it abolishes the ability of *dnaA recG tus rpo** cells to establish colonies at 42 °C and reduces amplification of the terminus area (Fig. 2b–d). The residual amplification observed depends on forks established at *oriC* (Fig. 2c and Supplementary Fig. 6), indicating that it is most probably the result of some replication through the hairpin (Supplementary Fig. 5a and Supplementary Discussion). Amplification is evident on both sides of the *tos* site (Fig. 2d), making it unlikely that it is due to activation of a dormant origin or of a hotspot for initiation at R-loops¹⁰.

But why should fork collisions trigger initiation? We have proposed that when replisome complexes meet, the DnaB helicase of one fork often displaces the nascent leading strand of the opposing fork, generating a 3' flap. DnaB would most probably collide with and dislodge the leading strand polymerase of the opposing fork to which the 3' end of the nascent leading strand is engaged (Fig. 3a and Supplementary Discussion)³. This flap would be degraded by 3' ssDNA exonucleases or converted by RecG to a 5' flap and subsequently

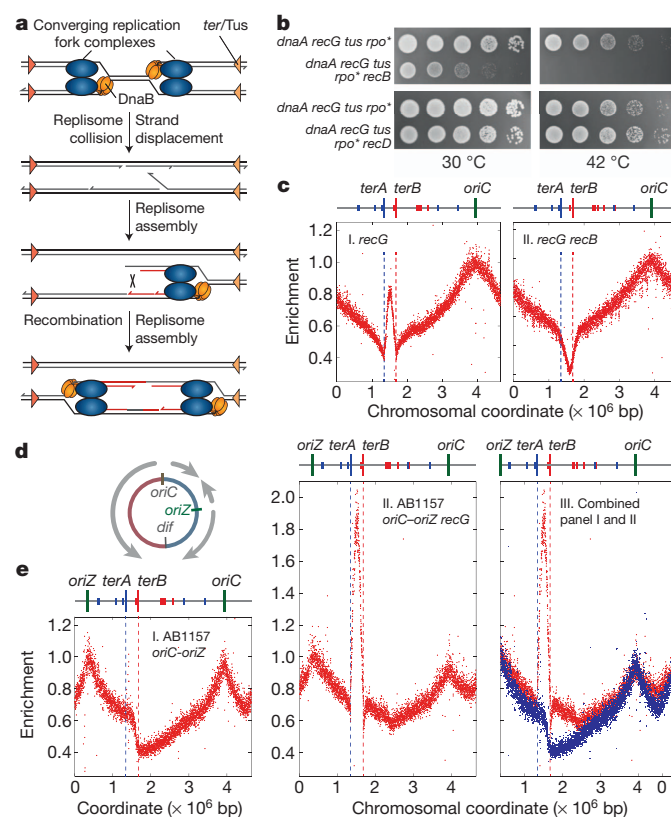


Figure 3 | Effect of RecBCD activity and *oriC* duplication on DnaA-independent replication. **a**, Model illustrating how collision of replication forks in the termination area might lead to the formation of two new divergent forks through PriA-mediated replisome assembly and RecBCD-mediated recombination. **b**, DnaA-independent growth. The strains were RCe268, RCe435 and RCe437. **c**, Effect of *recB* on the replication profile of *recG* cells. Sequencing templates were from N8226 and AM1581 cultured at 37 °C. **d**, Chromosome map of a double origin strain showing the positions of *oriC* and *oriZ* and illustrating where the forks assembled are likely to converge. **e**, Marker frequency analysis of chromosome replication in a double origin strain. Sequencing templates were from WX296 and RCe455 cultured at 37 °C. Chromosome coordinates in panel III are shifted to highlight the elevated marker frequency between *oriC* and *oriZ*.

degraded by a 5' exonuclease. Without RecG, a displaced 3' flap might persist, providing a substrate for PriA-mediated replisome assembly. Progression of this fork generates a duplex arm with a free end, which invades the homologous duplex behind the fork through RecBCD- and RecA-mediated recombination⁵. This creates a D-loop that PriA could also target to establish another fork moving in the opposite direction. Divergent replication initiated through recombination is consistent with our finding that growth of *dnaA recG tus rpo** cells at 42 °C is prevented by inactivating RecA, the RuvABC Holliday junction resolvase or the RecBCD recombinase (Fig. 3b and data not shown). Eliminating just the ExoV activity of RecBCD has no such effect (Fig. 3b, *recD* derivative⁵). Marker frequency analysis confirmed that inactivating RecB prevents over-replication of the terminus. Indeed, marker frequency is reduced in this region (Fig. 3c). This is a feature of cells lacking RecBCD as it is also seen in a *recB* single mutant (Supplementary Fig. 7a). BrdU incorporation confirmed that inactivating RecB reduces stable DNA replication in *recG* cells (Supplementary Fig. 7b). Taken together, these data indicate that without RecG, fork collisions often lead to the generation of dsDNA ends that are then targeted by RecBCD. The viability of *recG recB* double mutant cells¹⁴ excludes the possibility that these ends are the result of fork or chromosome breakage.

Profiling cells carrying two copies of *oriC* (Fig. 3d) showed that fork collisions trigger initiation of replication wherever forks meet. As expected¹⁷, both origins fire with equal efficiency on a population basis (Fig. 3e), leading to the termination of replication in two distinct zones, one in the area flanked by *ter* sites and a second in the shorter of the two intervals between the origins. The latter is shallower and less focused, which is to be expected given the absence of Tus/*ter* traps, whereas the former is deeper and more sharply defined. There is also a clear step between *terA* and *terB*, consistent with the clockwise fork from *oriZ* being blocked by Tus bound at *terB*, having reached this point before meeting the anticlockwise fork from *oriC*.

In the *recG* derivative, we observed a strong increase in amplification of the *terA-terB* interval (Fig. 3e and Supplementary Discussion). More notably, the termination region in the shorter interval is also noticeably shallower. This is precisely what one might expect if fork collisions were prone to trigger initiation. Without Tus/*ter* complexes to focus events, such initiation would lead to a wide region of amplification rather than to a distinct peak, as is seen when Tus is eliminated from a *recG* strain (cf. Figs 3c and 4a). Incomplete synchronization of origin firing¹⁷ would also reduce collisions in this

region. These data support the notion that replication fork collisions have the potential to trigger new replication wherever forks meet.

But how is replication maintained in *dnaA recG tus rpo** cells without *oriC* firing? Shifting the cells from 30 °C to 42 °C would not affect existing forks. When these collide at the terminus and trigger initiation, the new forks would proceed towards *oriC*. Any subsequent fork collisions would create an opportunity to repeat the cycle, potentially *ad infinitum*. With no Tus/*ter* complexes, collisions might become increasingly random. However, a broad peak of increased terminal markers is evident (Fig. 1c), suggesting perhaps that initiation is not random. Therefore, it is notable that the replication profile of *tus* cells in exponential phase is identical to wild type (Fig. 4a, b). Even in *recG tus* cells the termination area is still evident, although more broadly delineated (Fig. 4a). Thus, the sharp focusing of termination in wild-type cells is probably due to some factor in addition to Tus/*ter*. We suggest that the polarity of transcription impedes forks moving beyond the terminus towards *oriC*. Broadening of the terminus area in *tus rpo** cells supports this idea (Fig. 4b). However, *rpo** is unlikely to eliminate this problem entirely. Therefore, *dnaA recG tus rpo** cells with forks moving towards *oriC*, having duplicated the terminus, might be overrepresented, giving the marker frequency observed (Fig. 1c). Also, if forks moving towards *oriC* failed to complete duplication of the chromosome because of the conflicts with transcription, DNA from the affected (dead) cells would further bias the profile in favour of terminal markers.

Without means to control initiation, these cells lack the ability to coordinate chromosome replication with cell growth. New rounds of replication initiate only after replication fork collision, preventing the increased rate of growth and division made possible in wild-type cells by firing *oriC* before the previous round of replication has been completed³. So, what happens when growth ceases? Marker frequency analysis of *recG* cells grown to saturation showed an overrepresentation of the terminus (Fig. 4c), as if they had ceased growing with forks stalled at Tus/*ter*. An even broader elevation of terminal markers is observed in saturated *recG tus* cells. Coupled with its absence in *tus* cells (Supplementary Fig. 8), this persistence of forks suggests that once stationary-phase *dnaA recG tus rpo** cells are diluted in fresh growth medium, replication resumes at the stalled forks.

Taken together, the studies reported here demonstrate that RecG and 3' ssDNA exonucleases play critical roles in limiting re-replication of the already replicated DNA. It is notable that re-replication is blocked by mutations (*priA300*, *srgA1* and *ΔpriB*) that suppress the *recG* mutant phenotype¹⁸. This suggests that such replication has pathological consequences that destabilize the genome.

The single fork collision typical of the average *E. coli* cell cycle is a stark contrast to the multiple events occurring in eukaryotes. Recent studies have shown that the final stages of replication in eukaryotic cells have pathogenic potential^{19,20}. RecG is absent, but several studies have reflected on the ability of human and yeast helicases to remodel branched DNA structures in a manner reminiscent of RecG^{21–24}. Several ssDNA exonucleases have also been linked with replication, both nuclear and mitochondrial²⁵. The human mitochondrial nuclease MGME1, loss of which is associated with multi-systemic mitochondrial disease, has been shown to process displaced ssDNA flaps²⁵. So, as with *E. coli*, eukaryotic helicases and ssDNA exonucleases may well turn out to have important roles in making sure that replication is completed without triggering pathological events that destabilize the genome. Indeed, limiting the incidence of pathological events at termination may prove to be as crucial to genome stability as the prevention of re-initiation at replication origins².

METHODS SUMMARY

E. coli strains are described in Supplementary Table 1. The *dnaA46* allele encodes a thermosensitive DnaA protein that is inactive at 42 °C. Genetic manipulations, molecular procedures and methods for assessing growth and viability followed established protocols (see Supplementary Material). Replication profiling by

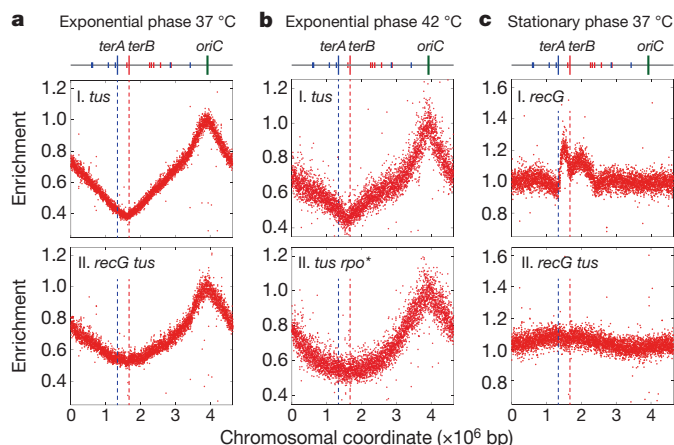


Figure 4 | Effect of *recG*, *tus* and growth phase on chromosome marker frequencies. **a, b,** Effect of *tus* and *rpo** on marker frequency in exponential-phase cells cultured at 37 °C (**a**) or 42 °C (**b**). Sequencing templates were from N8227, N7957, JJ1378 and RCe260. **c,** Effect of *recG* and *tus* on marker frequency in stationary-phase cells at 37 °C. Sequencing templates were from cultures incubated with vigorous aeration until well after no further increase in cell density was detectable. The strains were N6576 and N7957.

marker frequency analysis used AB SOLiD sequencing to measure sequence copy number. Enrichment of uniquely mapping sequence tags, in 1 kb windows, was calculated for an exponentially growing (replicating) sample relative to a non-replicating stationary-phase wild-type (MG1655) sample to correct for differences in read depth and to allow presentation of the data as a normalized marker frequency²⁶.

Full Methods and any associated references are available in the online version of the paper.

Received 22 February; accepted 17 May 2013.

Published online 28 July 2013.

1. Reyes-Lamothe, R., Wang, X. & Sherratt, D. *Escherichia coli* and its chromosome. *Trends Microbiol.* **16**, 238–245 (2008).
2. Diffley, J. F. Quality control in the initiation of eukaryotic DNA replication. *Phil. Trans. R. Soc. B* **366**, 3545–3553 (2011).
3. Rudolph, C. J., Upton, A. L., Briggs, G. S. & Lloyd, R. G. Is RecG a general guardian of the bacterial genome? *DNA Repair* **9**, 210–223 (2010).
4. Gabbai, C. B. & Marians, K. J. Recruitment to stalled replication forks of the PriA DNA helicase and replisome-loading activities is essential for survival. *DNA Repair* **9**, 202–209 (2010).
5. Kowalczykowski, S. C. Initiation of genetic recombination and recombination-dependent replication. *Trends Biochem. Sci.* **25**, 156–165 (2000).
6. Kim, N. & Jinks-Robertson, S. Transcription as a source of genome instability. *Nature Rev. Genet.* **13**, 204–214 (2012).
7. Paul, S., Million-Weaver, S., Chattopadhyay, S., Sokurenko, E. & Merrikh, H. Accelerated gene evolution through replication-transcription conflicts. *Nature* **495**, 512–515 (2013).
8. Rudolph, C. J., Dhillon, P., Moore, T. & Lloyd, R. G. Avoiding and resolving conflicts between DNA replication and transcription. *DNA Repair* **6**, 981–993 (2007).
9. Mott, M. L. & Berger, J. M. DNA replication initiation: mechanisms and regulation in bacteria. *Nature Rev. Microbiol.* **5**, 343–354 (2007).
10. Kogoma, T. Stable DNA replication: Interplay between DNA replication, homologous recombination, and transcription. *Microbiol. Mol. Biol. Rev.* **61**, 212–238 (1997).
11. Skovgaard, O., Bak, M., Lobner-Olesen, A. & Tommerup, N. Genome-wide detection of chromosomal rearrangements, indels, and mutations in circular chromosomes by short read sequencing. *Genome Res.* **21**, 1388–1393 (2011).
12. Trautinger, B. W., Jaktaji, R. P., Rusakova, E. & Lloyd, R. G. RNA polymerase modulators and DNA repair activities resolve conflicts between DNA replication and transcription. *Mol. Cell* **19**, 247–258 (2005).
13. Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M. E. & Nudler, E. Linking RNA polymerase backtracking to genome instability in *E. coli*. *Cell* **146**, 533–543 (2011).
14. Gregg, A. V., McGlynn, P., Jaktaji, R. P. & Lloyd, R. G. Direct rescue of stalled DNA replication forks via the combined action of PriA and RecG helicase activities. *Mol. Cell* **9**, 241–251 (2002).
15. Rudolph, C. J., Mahdi, A. A., Upton, A. L. & Lloyd, R. G. RecG protein and single-strand DNA exonucleases avoid cell lethality associated with PriA helicase activity in *Escherichia coli*. *Genetics* **186**, 473–492 (2010).
16. Cui, T. *et al.* *Escherichia coli* with a linear genome. *EMBO Rep.* **8**, 181–187 (2007).
17. Wang, X., Lesterlin, C., Reyes-Lamothe, R., Ball, G. & Sherratt, D. J. Replication and segregation of an *Escherichia coli* chromosome with two replication origins. *Proc. Natl Acad. Sci. USA* **108**, E243–E250 (2011).
18. Mahdi, A. A., Briggs, G. S. & Lloyd, R. G. Modulation of DNA damage tolerance in *Escherichia coli* recG and ruv strains by mutations affecting PriB, the ribosome and RNA polymerase. *Mol. Microbiol.* **86**, 675–691 (2012).
19. Fachinetti, D. *et al.* Replication termination at eukaryotic chromosomes is mediated by Top2 and occurs at genomic loci containing pausing elements. *Mol. Cell* **39**, 595–605 (2010).
20. Steinacher, R., Osman, F., Dalgaard, J. Z., Lorenz, A. & Whitby, M. C. The DNA helicase Pfh1 promotes fork merging at replication termination sites to ensure genome stability. *Genes Dev.* **26**, 594–602 (2012).
21. Ralf, C., Hickson, I. D. & Wu, L. The Bloom's syndrome helicase can promote the regression of a model replication fork. *J. Biol. Chem.* **281**, 22839–22846 (2006).
22. Whitby, M. C. The FANCM family of DNA helicases/translocases. *DNA Repair* **9**, 224–236 (2010).
23. Betous, R. *et al.* SMARCA1 catalyzes fork regression and Holliday junction migration to maintain genome stability during DNA replication. *Genes Dev.* **26**, 151–162 (2012).
24. Killen, M. W., Stults, D. M., Wilson, W. A. & Pierce, A. J. *Escherichia coli* RecG functionally suppresses human Bloom syndrome phenotypes. *BMC Mol. Biol.* **13**, 33 (2012).
25. Kornblum, C. *et al.* Loss-of-function mutations in MGME1 impair mtDNA replication and cause multisystemic mitochondrial disease. *Nature Genet.* **45**, 214–219 (2013).
26. Müller, C. A. & Nieduszynski, C. A. Conservation of replication timing reveals global and local regulation of replication origin activity. *Genome Res.* **22**, 1953–1962 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank T. Horiuchi and D. Sherratt for *E. coli* strains, A. Mahdi for help with DNA extractions, S. Malla and M. Blythe for deep sequencing, S. Demolli and D. Ivanova for control experiments, and C. Buckman and L. Harris for assistance. This work was supported by grants from the MRC (R.G.L., G0800970), the Leverhulme Trust (C.J.R.) and the BBSRC (C.A.N., BB/E023754/1).

Author Contributions C.J.R. and R.G.L. initiated and directed the project. C.J.R., A.L.U., A.S., C.A.N. and R.G.L. performed the experimental work. C.J.R., A.L.U., C.A.N. and R.G.L. analysed the data and wrote the paper.

Author Information Deep sequencing data have been deposited with NCBI Gene Expression Omnibus under accession number GSE41975. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.J.R. (christian.rudolph@brunel.ac.uk).

METHODS

Bacterial strains and general methods. The *E. coli* strains used in this study are derivatives of *E. coli* K-12 MG1655 unless stated otherwise (Supplementary Table 1). The *dnaA46* allele encodes a thermosensitive DnaA protein that is inactive at 42 °C. The *oriC* deletion allele tagged with a sequence encoding resistance to kanamycin (*oriC::kan*) was generated by the single-step gene replacement method²⁷. Luria and Burrous broth and 56/2 salts media have been cited elsewhere^{12,28}. Strain constructions, genetic manipulations and assays for bacterial and phage growth used standard microbiological materials and protocols, as described or cited^{18,29}. For assessing growth without DnaA, cultures of *dnaA46* constructs grown at 30 °C to an absorbance ($A_{650\text{ nm}}$) of 0.4 ($\sim 2 \times 10^8$ cells per millilitre) were diluted in tenfold steps from 10^{-1} to 10^{-5} before spotting 10 μ l samples of each dilution on LB agar. Duplicate plates were incubated at 30 °C and 42 °C. Pictures were taken after 24 h unless stated otherwise.

Replication profiling by marker frequency analysis. Samples from cultures of a strain grown to $A_{650\text{ nm}} \sim 1.2$ in LB broth were diluted 400-fold in fresh broth and incubated with vigorous aeration until $A_{650\text{ nm}}$ reached 0.4 at the temperature indicated. Samples from these exponential-phase cultures were frozen in liquid nitrogen at this point for subsequent DNA extraction. Incubation of the remaining culture was continued until several hours after the culture had saturated and showed no further increase in $A_{650\text{ nm}}$. A further sample (stationary phase) was frozen at this point. DNA was then extracted using a GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich). Marker frequency analysis used AB SOLiD sequencing to measure sequence copy number. Enrichment of uniquely mapping sequence tags, in 1 kb windows, was calculated for an exponentially growing (replicating) sample relative to a non-replicating stationary-phase wild-type sample to correct for differences in read depth across the genome and to allow presentation of the data as a marker frequency, as described previously²⁶.

BrdU labelling. BrdU labelling and detection by immunostaining was essentially as described³⁰. Cells were grown in 56/2 salts supplemented with 0.2% casamino acids and 0.32% glucose to $A_{650\text{ nm}} = 0.2$. The culture was split into various 2 ml aliquots, BrdU (Sigma) added to the first aliquot to 20 $\mu\text{g ml}^{-1}$ and the cultures shifted to 42 °C. At the times indicated, BrdU was added to one of the remaining aliquots. The aliquots were labelled with BrdU for 8 min, pelleted and re-suspended in 85 μ l TEE buffer (10 mM Tris•HCl, 10 mM EGTA, 100 mM EDTA, pH 8.0), containing 0.05% lauroylsarcosine and 0.5% SDS. Eighty-five microlitres of liquid 1.4% low melting point agarose was added and the mixture solidified in a disposable plug former (Bio-Rad) at 4 °C. Plugs were treated with 10 mg ml^{-1} lysozyme in 3 ml TEE buffer containing 0.05% lauroylsarcosine and 0.5% SDS for 2 h at 37 °C and then at 52 °C overnight with 5 mg ml^{-1} proteinase K in 3 ml TEE containing 1% SDS. Plugs were washed in TEE for 30 min at 37 °C, treated with 1 mM phenylmethane sulphonyl fluoride (freshly prepared as 100 mM stock solution in methanol) in fresh TEE for 1 h at 37 °C, washed in fresh

TEE for 30 min at 37 °C and finally in $0.1 \times$ TEE for 30 min at 37 °C. The plugs were subsequently transferred into 300 μ l restriction enzyme buffer and incubated for 30 min at room temperature, the buffer changed and 25 units of NotI (NEB) added. Chromosomal DNA was digested overnight and the fragments separated on a 0.8% agarose gel (Bio-Rad pulse field certified agarose) in $0.5 \times$ TBE using a CHEF Mapper PFGE system (Bio-Rad), running with a gradient voltage of 6 V cm^{-1} , an included angle of 120°, and initial and final switch times of 1.65 and 32.45 s, respectively, with a run time of 20 h at 14 °C. DNA was transferred to a Hybond-N+ Membrane (GE Healthcare) by alkaline vacuum transfer and ultra-violet crosslinked (120 mJ cm^{-2}). Blocking was achieved with TBS Tween (50 mM Tris•HCl, 150 mM NaCl, pH 8.0, 0.5% Tween 20) containing milk powder (2%). After blocking the membrane was incubated for 2 h in the presence of mouse anti-BrdU antibody (Santa Cruz) and diluted 1:5,000 in TBS Tween. Horseradish-peroxidase-conjugated secondary antibody (goat anti-mouse, Bio-Rad) was used at a dilution in TBS Tween of 1:10,000 for 1.5 h. The membrane was incubated with ECL Plus Western Blotting Detection Reagents (GE Healthcare) and the signal visualized either by exposure to X-Omat UV Plus film (Kodak) or by the ChemiDoc chemiluminescence detection system (Bio-Rad).

Chromosome linearization. Linearization of the *E. coli* chromosome was achieved using the bacteriophage N15 telomere generating system (Supplementary Fig. 5). N15 is a temperate *E. coli* phage, which does not integrate into the chromosome during lysogenization³¹. Instead, the N15 telomerase, TelN, cleaves and processes a specific phage DNA sequence called *tos*, generating a linear chromosome with hairpin ends (Supplementary Fig. 5). When replication forks reach these structures, the nascent leading and lagging strands are sealed, forming a chromosomal dimer. This restores two *tos* sites, which are immediately re-processed by TelN^{16,31}, thus allowing segregation of two linear chromosomes. For our studies, we exploited *E. coli* derivatives that carried a *tos* site inserted near the *dif* locus, close to *terC*¹⁶. These were lysogenized with phage N15 to linearize the chromosome (Supplementary Fig. 5). Preparation of phage N15 stocks and *E. coli* lysogens followed methods and protocols described for phage λ ²⁹.

27. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
28. McGlynn, P. & Lloyd, R. G. Modulation of RNA polymerase by (p)ppGpp reveals a RecG-dependent mechanism for replication fork progression. *Cell* **101**, 35–45 (2000).
29. Miller, J. H. *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, 1972).
30. Rudolph, C. J., Upton, A. L. & Lloyd, R. G. Replication fork collisions cause pathological chromosomal amplification in cells lacking RecG DNA translocase. *Mol. Microbiol.* **74**, 940–955 (2009).
31. Rybchin, V. N. & Svarchevsky, A. N. The plasmid prophage N15: a linear DNA with covalently closed ends. *Mol. Microbiol.* **33**, 895–903 (1999).

CORRIGENDUM

doi:10.1038/nature12370

Corrigendum: Inhibition dominates sensory responses in awake cortex

B. Haider, M. Häusser & M. Carandini

Nature **493**, 97–100 (2013); doi:10.1038/nature11655

In this Letter, there were two typographical errors in both the Methods Summary and the full Methods. The concentration of urethane should be 1.5 g per kg of body weight, not 1.5 mg. In addition, the intracellular solutions contained 0.3 mM Na₃GTP, not Na₂ATP. These errors have been corrected in the HTML and PDF versions of the manuscript. There were also errors in Supplementary Fig. 8, in which we inadvertently combined data from the wrong trial types to assess current–voltage (*I*–*V*) linearity. This has been corrected in the Supplementary Information of the original manuscript, and does not affect the results or conclusions.

ERRATUM

doi:10.1038/nature12431

Erratum: Masses of exotic calcium isotopes pin down nuclear forces

F. Wienholtz, D. Beck, K. Blaum, Ch. Borgmann, M. Breitenfeldt, R. B. Cakirli, S. George, F. Herfurth, J. D. Holt, M. Kowalska, S. Kreim, D. Lunney, V. Manea, J. Menéndez, D. Neidherr, M. Rosenbusch, L. Schweikhard, A. Schwenk, J. Simonis, J. Stanja, R. N. Wolf & K. Zuber

Nature **498**, 346–349 (2013); doi:10.1038/nature12226

In Fig. 3b of this Letter, the *y* axis numbering should be identical to that of Fig. 3a, and start at 2 MeV and go up to 25 MeV. Instead, the numbering was inadvertently shifted upwards, which made the data shown in Fig. 3b appear to have higher values by 2 MeV with respect to their actual ones. Figure 3 has been corrected in the HTML and PDF of the original manuscript.

CORRIGENDUM

doi:10.1038/nature12418

Corrigendum: Sema3A regulates bone-mass accrual through sensory innervations

Toru Fukuda, Shu Takeda, Ren Xu, Hiroki Ochi, Satoko Sunamura, Tsuyoshi Sato, Shinsuke Shibata, Yutaka Yoshida, Zirong Gu, Ayako Kimura, Chengshan Ma, Cheng Xu, Waka Bando, Koji Fujita, Kenichi Shinomiya, Takashi Hirai, Yoshinori Asou, Mitsuhiro Enomoto, Hideyuki Okano, Atsushi Okawa & Hiroshi Itoh

Nature **497**, 490–493 (2013); doi:10.1038/nature12115

In this Letter the left panels of Fig. 2a show images of a femur from a semaphorin 3A-knockout (*Sema3a*^{col1}^{−/−}) mouse, rather than from a littermate *Sema3a*^{fl/fl} mouse. Figure 1 shows the correct panels. This error does not affect our conclusions and the legend of Fig. 2 is correct.

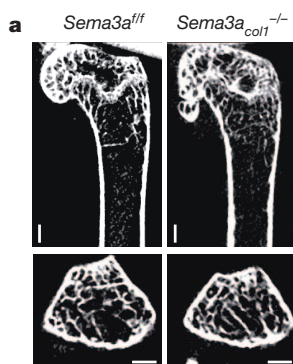


Figure 1 | This figure shows the microcomputed tomography panels of Fig. 2a with the left panels corrected.

CAREERS

TURNING POINT Researcher looks to nature to develop innovative materials **p.615**

MEDIA Surveys suggest scientists wary of pre-publication media interaction **p.615**

NATUREJOBS For the latest career listings and advice www.naturejobs.com

DAVID SALES BATISTA/SHUTTERSTOCK



COLUMN

Biomedical burnout

Stress, long hours and low morale threaten to scar the activities and careers of US life-sciences researchers, argue **Warren Holleman** and **Ellen R. Gritz**.

Studies suggest that burnout among medical doctors has reached epidemic proportions around the world, accompanied by alarmingly high levels of clinical depression, suicidal thoughts, job dissatisfaction and unhappiness with work-life balance¹. The data are so compelling that some health organizations and physician groups are exploring ways to tackle these problems².

Almost nothing is known, however, about stress, burnout and morale among biomedical scientists. To take a closer look, we interviewed the chairs of the science departments

at MD Anderson Cancer Center in Houston, Texas, between July 2011 and February 2012. We asked them to assess changes in faculty morale, to identify major sources of stress and to discuss their feelings of optimism or pessimism about the future. We found that many faculty members are struggling in the face of funding pressures, bureaucracy, administrative burdens and faculty-administration conflict.

Without realizing it, we had captured the mood of an institution about to experience a difficult period that accentuated such pressures. A few months after we collected our results,

MD Anderson leadership came under fire, and the centre's faculty senate conducted a separate survey to assess faculty concerns (see go.nature.com/jcmgv2). The survey results themselves, which suggested high clinical workload, dismay over the departure of valued leaders and displeasure with top leadership over charges of nepotism and conflicts of interest, further upset faculty members.

In addition to conducting our own survey, we have talked to people at other institutions, and examined relevant publications^{3,4}. As a result, we conclude that the discontent at ►

► MD Anderson is emblematic of distressingly low morale at centres around the country. The effects of this year's federal budget sequester have only added to the gloom (see *Nature* **498**, 527–528; 2013). Ours is a pilot effort that we hope will be the basis for wider exploration and study into burnout among biomedical scientists — already, we have had preliminary discussions about a multi-centre study.

Financial and other pressures are certainly not unprecedented at academic institutions. But in recent years, our findings suggest, they have intensified and exacerbated each other, making research environments particularly stressful.

A MEASURE OF MORALE

In our survey, we approached 21 department chairs. We interviewed 19; three focused on conditions unique to their own departments, but the remaining 16 shared their observations, opinions and feelings about morale in general. The responses reflect pressing issues at one institution, but suggest struggles throughout the biomedical-research community.

Most of the chairs said that the morale of faculty scientists has worsened in recent years. Seven said that it had worsened significantly. As one interviewee put it, “Many faculty are deflated, unsettled and depressed. There is a sense of hopelessness; they’ve given up. There is some resentment; they’ve spent a long time establishing their careers and now there doesn’t seem to be a way to continue doing what they like to do.” Another alluded to the uncertain future of young investigators: “When I was a postdoc, the sense was that if you’re good, you’ll find a job. I can’t say that any more.”

The main stressors seem similar across different departments. Not surprisingly, funding was a big one. Several of the chairs described academic scientists as caught in a “perfect storm”: at a time when funding rates at the US National Institutes of Health have dropped drastically, some institutions are requiring faculty members to raise higher percentages of their salaries through grants. Productive scientists at middle and even senior levels are struggling to keep their careers afloat. “These are solid scientists, not marginal scientists. I don’t remember it ever being like that in the past. This hurts morale,” said one respondent.

The department chairs felt that institutions see ever-increasing productivity as a major goal. Executive leaders expect faculty members to seek and obtain more grant funding continuously, achieve higher-quality publications (as measured by metrics such as journal impact factors) and generally meet higher standards for academic excellence in areas such as teaching and collaboration. Although these are admirable goals, respondents noted that constantly raising the bar for high-functioning faculty members intensifies levels of comparison and implicit competition.

There is also the pressure to be continually

innovative in terms of research proposals, findings, publications and, in some cases, commercialization potential. Several respondents expressed concern that this relentless pressure might cause stress and burnout, and one said that department heads worry about extreme responses such as suicide or research misconduct aimed at gaming the system.

Some chairs discussed a downward spiral in which a scientist loses funding, and as a result has to reduce lab space and personnel, which in turn makes it more difficult to obtain funding. This dooms the scientist’s chances as an independent investigator. Even when a career is salvaged, the researcher often suffers ongoing emotional strain as a result of losing valuable time and resources, and experiences guilt and shame about the adverse impact on the lives and careers of lab personnel.

Administrative duties are another stressor. Ten interviewees complained that reports,

“Financial and other pressures have intensified and exacerbated each another, making research environments stressful.”

paperwork, personnel evaluations, grant procedures, training requirements, business meetings and daily e-mail minutiae have become much more cumbersome than in the past, distracting scientists from research and sapping energy, creativity and productivity. “There are more administrative and reporting demands, to the point where they seem repetitive, overlapping and always on a high-demand time schedule,” said one respondent. Another said that they “spend much more time jumping through hoops”.

A NEW APPROACH

This all paints a rather bleak picture. But we think that MD Anderson has made inroads towards addressing the problem as an institution, even in light of recent challenges. Twelve years ago, in response to the suicide of a colleague, a group of concerned faculty members and executive leaders formed a Faculty Health Committee (with E.R.G. as the founding chair) to develop a crisis-response protocol and to consider other ways to preserve and promote faculty welfare. The committee developed the Faculty Health & Well-Being programme, of which W.H. is the director.

Many of the programme’s activities, such as lectures and departmental seminars, are educational. Others are designed to enhance peer support as a buffer against stress and burnout. Several are outside the realm of academic medicine: social gatherings, dance lessons, parenting support groups, opera performances, a faculty art show, meditation, yoga and t’ai chi. We launched a Faculty Assistance Program to enhance access to mental health care. We also facilitated panel discussions on work–life balance, dialogues with executive leaders and a three-day working conference on faculty

health and well-being, the result of which was a book, *Faculty Health in Academic Medicine: Physicians, Scientists, and the Pressures of Success*⁵. Our survey results show that we still have work to do, but we have a framework in place to address problems and offer faculty support.

Our faculty senate collects input from MD Anderson scientists and maintains communication with institutional leaders through multiple channels, raising morale-related issues and proposing solutions. In the case of concerns and controversies about the centre’s financial decision-making, clinical operations and conflict-of-interest policies in the past few years, the faculty senate organized meetings with executive leaders to improve communication, trust, transparency and shared decision-making.

At such meetings, leaders learned about the sources of faculty stress, and helped to brainstorm ways to mitigate them. We have implemented initiatives to reduce paperwork through a task force and an information-technology overhaul. And we have bolstered institutional mechanisms to provide extra financial support through bridge funding, seed funding and departmental-chair funds.

Through exchanges with leaders, faculty members have learned about the financial and regulatory realities of an institution. Progress has sometimes been slow — and we cannot remove every obstacle. But the more dialogue we have and the more effectively we collaborate, the better faculty members will understand challenges, appreciate efforts, engage in the process of negotiating change and finding compromises, and stay hopeful about future progress.

Could other institutions implement similar strategies? Yes, with sustained support from leaders. It sometimes takes a dramatic event, such as a financial or organizational crisis, to overcome institutional inertia and to bring executive and faculty leaders together to address their common interests. It also helps to have data. The MD Anderson faculty senate’s morale survey has documented the need for the Faculty Health & Well-Being programme, and has provided our leadership with incentives and directions for cultural change.

These efforts are just a start, but we think a good one. It is crucial that we try to make a difference — now more than ever. ■

Warren Holleman is professor and **Ellen R. Gritz** is chair of the department of behavioural science at the University of Texas MD Anderson Cancer Center in Houston.

1. Shanafelt, T. D. *et al.* *Arch. Intern. Med.* **172**, 1377–1385 (2012).
2. Kuerer, H. M. & Holleman, W. L. *Ann. Surg.* **255**, 634–636 (2012).
3. National Research Council *Research Universities and the Future of America* (National Academies Press, 2012).
4. Ginsberg, B. *The Fall of the Faculty* (Oxford Univ. Press, 2011).
5. Cole, T. R., Goodrich, T. J. & Gritz, E. R. (eds) *Faculty Health in Academic Medicine* (Humana, 2009).

TURNING POINT

Angela Belcher

MIT Biomolecular materials researcher Angela Belcher looks to nature for inspiration; her observations have helped her to devise materials for clean fuels and efficient batteries. In June, Belcher, of the Massachusetts Institute of Technology (MIT) in Cambridge, was named the recipient of a US\$500,000 Lemelson-MIT Prize honouring mid-career scientists who make an impact on society.

You had an unusual undergraduate path.

How did it influence your career?

I did my bachelor's degree in creative studies, which allowed me to design my own major at the University of California, Santa Barbara (UCSB). It was one of the best decisions I ever made. I could take risks, and it allowed me to put different topics together. I worked in plant molecular biology, physics and chemistry labs and did ecology research. Later, as a doctoral student, I worked with a physicist, a molecular biologist and an inorganic chemist. It was like a playground for multidisciplinary science.

How did you explore your multidisciplinary interests after graduating?

I isolated proteins involved in growing abalone shell and became fascinated that a soluble protein made by an animal could generate this material. I wanted to apply these concepts to electronics, but I didn't have the background. I did a postdoc with Evelyn Hu, then head of an electrical-engineering lab at UCSB. It was life-changing because I was able to start developing an interface between biological materials, genetics and semiconductor materials.

How did you find your feet in academia?

I explored multiple opportunities: I applied for industry jobs and a postdoc fellowship in entrepreneurship. I also started applying for faculty positions for practice, and got an offer at the University of Texas at Austin. I took the job and started working on selecting evolving organisms, such as viruses, to assemble new materials. My grant applications were not well received — until the first paper out of my lab was in *Nature* (S. R. Whaley *et al.* *Nature* **405**, 665–668; 2000), showing it could work. That led to the patents and the venture capital to start my first company, Cambrios Technologies in Sunnyvale, California, which develops electronic materials for touch screens.

You were recruited by two MIT departments jointly. Do you recommend that approach?

I don't advise people to pursue dual appointments as young faculty, because you have



commitments in both. It worked out well for me because it fit my research needs and teaching interests, and I had great faculty support. But I suggest that young people choose one department to join and be affiliated with others. That said, a couple of years ago I added a third faculty-level commitment, at the David H. Koch Institute for Integrative Cancer Research at MIT. I was worried that I would not have enough to contribute, but that has not been the case. I discuss topics with a mixture of engineers and cancer biologists.

How has being at MIT helped to foster your entrepreneurial goals?

I have had a lot of mentorship from faculty members to learn how to develop a product, and I have started a second company, Siluria Technologies in San Francisco, California, focused on converting methane to liquid transportation fuel. I tell young faculty members that to find a good academic home, they should decide what they want in terms of lab size, entrepreneurship and lifestyle, and should see if people at the institutions they are interested in are accomplishing those things.

What are the parallels between forming a company and nurturing a productive lab?

Both are founded on ideas. You just have to translate those ideas for a different purpose in business. The hard part is figuring out how to run a business, which scientists do not get much training in. Getting that on-the-job training — for example, learning how to interact with customers — is a humbling experience. One parallel is that hiring really smart people is key. I hire people who are passionate, smart and creative, and aren't afraid to admit they don't know everything. ■

INTERVIEW BY VIRGINIA GEWIN

MEDIA

Results kept quiet

Many scientists are wary of discussing unpublished results with journalists, suggests a meta-analysis of survey data (H. P. Peters *Proc. Natl Acad. Sci. USA* <http://doi.org/nhz>; 2013). Hans Peter Peters, a social scientist at the Jülich Research Centre in Germany, examined published and unpublished data that he and his colleagues collected in surveys of US neuroscientists and German scientists. A little more than half of US and German neuroscientists, and 44% of German scientists across all fields, agreed that acceptance of a paper by a journal is threatened if the findings have already been revealed in the mass media. Peters notes that most scientists make a sharp distinction between how they communicate among themselves and how they talk to the media and the public.

FUNDING

NSF to support forensics

The US National Science Foundation (NSF) is seeking proposals for basic research in forensic science, in an effort to improve rigour and standards. The move is partly a result of a 2009 US National Research Council report that called for the NSF to support forensics research, and for scientists and medical officers in the field to be certified, says Mark Weiss, division director for behavioural and cognitive science at the NSF in Arlington, Virginia. "If you've got an idea, we want to hear about it," says Weiss, who adds that forensics research is considered a national and legislative priority. He encourages interested researchers to contact programme directors in any relevant NSF directorate.

ADVOCACY

Postdoc representative

The US National Postdoctoral Association (NPA) in Washington DC has appointed a new executive director. Belinda Huang, a consultant and former administrator at the University of Pennsylvania in Philadelphia, says that she will reach out to lawmakers and to firms in biomedicine and other fields to boost the NPA's visibility and to promote the postdoc's role in US research. She will also build fund-raising efforts and apply for more grants. Huang succeeds Cathie Johnson Phillips, who had been NPA executive director since 2008 and had led many initiatives, including a call for increases to postdoc stipends.

THE SPEED OF DARK ENERGY

Faster than a speeding photon.

BY JEFF HECHT

The university's Innovation Networking Conference was going poorly, like the rest of Boris's life. His postdoc was running out, and with no job in the offing, he faced a one-way flight back to Novosibirsk. His eight papers in theoretical astrophysics would earn him no more in Russia than in America, but at least he could sleep on his father's couch. Boris was brooding when he noticed a portly, well-dressed man reading his poster.

"May I help you?" Boris asked, as Professor Liu had told him to.

The man nodded. "This dark-energy stuff in your poster is what makes the Universe expand faster, right? How fast does it go?"

Professor Liu had said not to answer simple questions with abstruse mathematics. "Yes, it permeates space, but what velocity do you mean?"

"Can it move faster than light?"

That puzzled Boris. "Why do you want to know?"

The man looked into Boris's eyes, and his voice dropped to a whisper. "Because I need something faster than light for a consulting project."

"Well," said Boris, pausing to search for the right words in the hope of somehow landing a job, "I don't know why it couldn't move faster than light." His grandfather liked to tell of a colleague who told a Brezhnev-era general that he didn't know why he couldn't build a 10-megawatt laser weapon. That statement had earned the man a good job and a dacha, and because his statement was literally true and because Brezhnev wasn't Stalin, nobody had got shot when it didn't work after a dozen years.

"Stop by my office in the morning," the man said, handing Boris a card identifying him as Andrew Harrison Harding, professor of economics.

"Nanoseconds count," began Professor Harding. "What do you know about high-frequency automated trading?"

Boris shrugged.

"It's how the stock market works now. Computers buy and sell stock to each other, and we hope our computers outsmart

theirs. The best way to outsmart them is to get information faster, so submarine fibre-optic cables run the shortest routes. One goes through the Arctic Ocean from Tokyo to London. Now our competitors are laying a hollow-core fibre along that route, and

interactions between dark energy and ordinary matter, then converted them into code for models to run on the quantum computer and his desktop supercomputer.

He tested the models for weeks, pored over the results, rewrote the code, and ran

the models again. The results mapped the distribution of dark energy relative to matter, so he calculated the rate of change and converted that into velocity. His first velocity results were close to zero, so he changed his reference frame, and found that the velocities were a little higher when he computed over more space. Impatient, he increased the scale by ten orders of magnitude and left the model running over the weekend, hoping that the computing strain wouldn't melt anything.

His desktop's fans howled like a wind tunnel on Monday morning when he asked it to display the results on a wall screen. The pattern looked familiar, but when he checked the colour scale, he was amazed to find that the yellow midpoint was the speed of light, and vast areas of green and blue showed superluminal velocities.

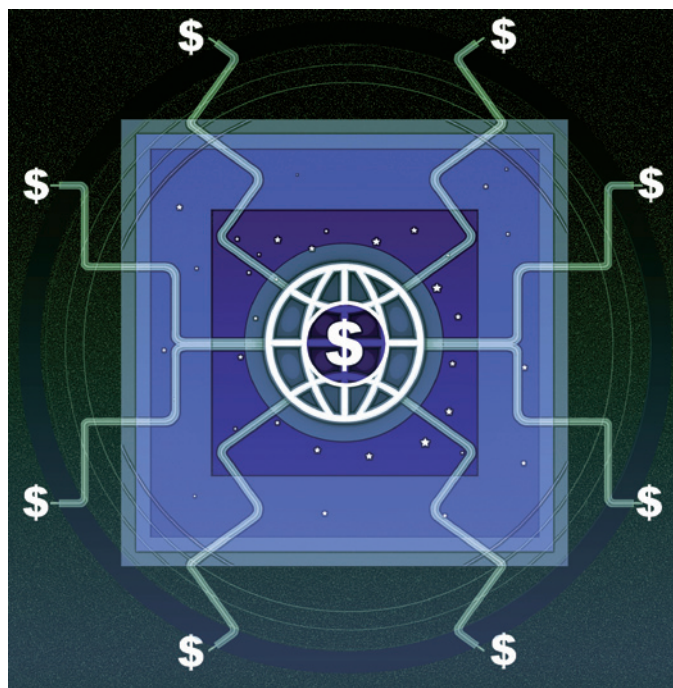
Professor Harding arrived minutes after Boris called him, and stared at the wall screen. "We're going to be rich!" he said. "Route the cable like this," he said, sweeping his arm on an arc.

"No cable," Boris said. "Dark energy permeates everything, so it goes through Earth."

"Better yet," said the professor. "Write this up and we'll patent it. Investors will be pounding on the door. I'll double your salary."

After Professor Harding left, Boris fiddled with the display. Curious how much space the model covered, he asked the supercomputer to show the linear scale. Small words appeared near the bottom: "Width of region shown, 10^{12} light years." For a few minutes, he felt very proud of himself for modelling beyond the edge of the observable Universe. Then he realized that the only superluminal velocities were beyond the edge. He was only a few orders of magnitude off. Maybe he could forestall the bad news as long as his grandfather's friend had. ■

Jeff Hecht is Boston correspondent for *New Scientist* and a contributing editor to *Laser Focus World*.



because light goes faster through the empty core than through a solid glass fibre, it will be faster than our cable. So we need to send data faster than light. Can dark energy do it?"

"It may in some places," Boris answered, not mentioning that those places were beyond the observable Universe. "It would take time to work out the details."

"I can pay you good money to do that."

Boris could invent equations, not cable systems, but this was no time to mention that. "I will need resources," he said.

"I have office space and access to the university's quantum computer." The professor paused. "You can start at \$2,000 a week on Monday."

Boris smiled. Professor Liu would be amazed.

Boris enjoyed theory. He filled bound quadrille-ruled notebooks with symbols and equations. He wrote mathematical descriptions of the

➔ NATURE.COM

Follow Futures:

🐦 @NatureFutures

📘 go.nature.com/mtoodm